

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2002, 08:20:28 ; Search time 1832.12 Seconds

(without alignments)  
11719.007 Million cell updates/sec

Title: US-09-664-186-4

Perfect score: 1026

Sequence: 1 gtgaagaagaacaaacatt.....ccctcaaggagcctcctga 1026

Scoring table: IDENTITY\_NUC

Searched: 1797656 seqs, 10463268293 residues

al number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*

1: gb.ba:\*

2: gb.htg:\*

3: gb.in:\*

4: gb.om:\*

5: gb.ov:\*

6: gb.pat:\*

7: gb.ph:\*

8: gb.pl:\*

9: gb.pr:\*

10: gb.to:\*

11: gb.sts:\*

12: gb.sy:\*

13: gb.un:\*

14: gb.vl:\*

15: em.ba:\*

16: em.fun:\*

17: em.hum:\*

18: em.in:\*

19: em.mu:\*

20: em.om:\*

21: em.or:\*

22: em.ov:\*

23: em.pat:\*

24: em.ph:\*

25: em.pl:\*

26: em.ro:\*

27: em.sts:\*

28: em.un:\*

29: em.vl:\*

30: em.htg.hum:\*

31: em.htg.in:\*

32: em.htg.inv:\*

33: em.htg.other:\*

34: em.htg.in:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result Query Score Match Length DB ID Description

Result	Query	Score	Match	Length	DB	ID	Description
1	1026	100.0	1026	6	AR139424		AR139424 Sequence
2	1026	100.0	1600	1	U89376		U89376 Thermus sp.
3	1026	100.0	5849	1	AF013571		AF013571 Thermus s
4	1026	100.0	5849	6	AR139426		AR139426 Sequence
5	59.8	5.8	38962	1	SC2E1		AL023797 Streptomy
6	56.8	5.5	13347	1	SCC42		AL135692 Streptomy
7	54.6	5.3	35576	1	SC2D46		AL191406 Streptomy
8	53.6	5.2	41622	1	SCD25		AL118514 Streptomy
9	53.2	5.2	100772	2	AC105260		AC105260 Oriza sat
10	53.2	5.2	167049	2	AP004746		AP004746 Oriza sat
11	52.6	5.1	7459	1	SC8A2		AL445327 Streptomy
12	52.6	5.1	34983	1	SCM11		AL133278 Streptomy
13	52.4	5.1	41347	1	SC4A10		AL109663 Streptomy
14	52.4	5.1	113193	1	AF357202		AF357202 Streptomy
15	52.2	5.1	188050	1	AL646072		AL646072 Ralstonia
16	51.8	5.0	19372	1	AE004443		AE004443 Pseudomon
17	51.6	5.0	1815	6	AX148314		AX148314 Sequence
18	51.6	5.0	2498	8	ZM095953		U95953 Zea mays vl
19	51.2	5.0	615	6	E04306		E04306 DNA encodin
20	51.2	5.0	915	1	TRHSUPDIS		D13387 Thermus aqu
21	51.2	5.0	6955	1	TRHILVB		D84646 Thermus aqu
22	50.8	4.9	23458	1	SCR22		AL355832 Streptomy
23	50.6	4.9	30941	1	SC4C6		AL079335 Streptomy
24	50.6	4.9	33810	1	SCB6		AL353832 Streptomy
25	50.4	4.9	10790	1	AE001973		AE001973 Deinococc
26	50.4	4.9	21220	1	SC1E6		AL033505 Streptomy
27	50.4	4.9	125020	9	AF429315		AF429315 Homo sapi
28	50.2	4.9	24397	1	AE004673		AE004673 Pseudomon
29	50.2	4.9	41906	1	SC5C7		AL031515 Streptomy
30	50.2	4.9	123980	2	AC103551		AC103551 Oriza sat
31	50	4.9	11187	1	AE001908		AE001908 Deinococc
32	50	4.9	39314	1	SGR300302		AJ300302 Streptomy
33	49.8	4.9	173356	2	AP004047		AP004047 Oriza sat
34	49.6	4.8	11308	1	SC9P2		AL035559 Streptomy
35	49.6	4.8	13814	1	AE007092		AE007092 Mycobacte
36	49.6	4.8	63033	1	MTV008		AL021246 Mycobacte
37	49.2	4.8	11790	1	SCC80		AL442143 Streptomy
38	49.2	4.8	13571	1	AE004982		AE004982 Halobacte
39	49.2	4.8	40104	1	SCR11		AL132662 Streptomy
40	49	4.8	1229	1	TC1DH		X04519 Thermus cal
41	49	4.8	12277	1	AF340166		AF340166 Streptomy
42	49	4.8	33320	1	SC9B10		AL009204 Streptomy
43	49	4.8	213050	1	AL646067		AL646067 Ralstonia
44	48.6	4.7	154746	14	HSV2HG52		Z86099 Herpes simp
45	48.4	4.7	2184	1	AF170451		AF170451 Rhizobium

## ALIGNMENTS

RESULT 1

LOCUS AR139424 1026 bp DNA

DEFINITION Sequence 4 from patent US 6207377.

ACCESSION AR139424

VERSION AR139424.1 GI:14481920

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1026)

AUTHORS Wayne,J. and Xu,S.

TITLE Method for construction of thermus-E. coli shuttle vectors and identification of two Thermus plasmid replication origins

JOURNAL Patent: US 6207377-A 4 27-MAR-2001;

FEATURES

source Location/Qualifiers

1..1026

/organism="unknown"

BASE COUNT 221 a 323 c 285 g 197 t

Query Match 100.0%; Score 1026; DB 6; Length 1026;



Db 344 GTGAAGAAAGAAAAACCTCTTTGAGAGCTTTACGAGGCTTTAGAGAAACCCAGAC 403  
 QY 61 aacacccagtcgcactaagggtgcagatagggtgcagagagcttctcttgccaccgac 120  
 Db 404 AACACGATGCGACTAGAGGGGTCAAGATAGGGGGTCAAGAGGACCTTCTTGGCCACACGAC 463  
 QY 121 cccctccagatggaggtgcccgaataatcgctccggaagggtcttaacatacaaaaagag 180  
 Db 464 CCCCCCAGATGAGAGGTGCGGAAATCGCTCGGAAGGGCTTTACATACCAAAAAAGAG 523  
 QY 181 gcaatagatgtgttaaccggagaagaacatgagagcttctcttctctgttgggccc 240  
 Db 524 GCACTTGGATGCTTATCCGAGAAAGACCATGAGGCTTCTTCTCTCTGTGTGGGGCC 583  
 QY 241 cccctatacaccacagctgaaacccccggtgggaatgataagacggtccaggaagg 300  
 Db 584 CCCCCCTATACACAGCTGAACCCCCGTTGGGAATGATATCAAGCGCTCCAGAGAGGG 643  
 QY 301 cctagaagaagcttctgaaactctccaaagagatltgcccgtccacacatccctacggaac 360  
 Db 644 CCTGAGAGCTTCTGGAACCTCTCCAGAGATTGCCCGCTCCACCATCCCTACGGGCAAC 703  
 QY 361 cggagagctctggaagagagtgaggagagtgctgtctcaatgttccccctgagatgtggcc 420  
 Db 704 CGGAGACTCTGAGAGAGGTGGGACGGTCTCTCATGTGTCCTCCCTGGAGATGTTGGCC 763  
 QY 421 ctaaaccttgggtgtcaaccgagacggtctcacgcccgtgagaagaaggtccttgaaaaaag 480  
 Db 764 CTCAACTGGGGGTGACCCGGCAGACCGTCCAGCGCTCGAAGAAAGTCTTGGAAAAAG 823  
 QY 481 ggcctgtgtccacccgaagcttcttcaacaaacggtcaaacggggagcgcgagccatcgac 540  
 Db 824 GGCCTGTGTGCGCACCGACGCTCTTCAACCAACCGTCAACGGGAGCGCGCGCATCGCG 883  
 QY 541 acccttgggcgctgcgagctgagccagggaaaagcagagctcctctgaagataatc 600  
 Db 884 ACCCTTGGGCGCTCGCGGTGAGAGCCAGGAAAGCCAGGCTCAACCTGAGCAGCTACATC 943  
 QY 601 taaccttgagaagaactcgccctagaacatgagccaaagcggtgctctccttcaactggtc 660  
 Db 944 TACCCCTGGAGGAACCTCGCCCTAGACATGCGCAACGGGTGCTCTCTCAACTGGGTC 1003  
 QY 661 aaggcctaccagaagcaggaatcgcccccacaccttgaagctgtgctctctgagctcag 720  
 Db 1004 AAGGCTTACCAGAACCAAGGAATCGCCCCACCCCTGAGCTGTGCTCTCTGGGCTCAG 1063  
 QY 721 gggaaaaaggtgattgccaacaacaaagacggtgagcttgaacttggccttcaatcctggtc 780  
 Db 1064 GGGAAAAAGGTGATGCCCAACCAAGACCGTGGCGGTGAGCTGGGCTCATCTGGGTC 1123  
 QY 781 ctcccccagatggagcggttccaaactccgcccctatacaacctcatltgctacatc 840  
 Db 1124 CTCCCGAAGGTGAGGCTTCAAAACCTCGCGCTTATACCTCATCTCTAGTACAT 1183  
 QY 841 gccgcatctcctagatgacgctgcttcaagaagcttctatgagagcttctgtgagctgtg 900  
 Db 1184 GCGGATCTCTAGATGACCGCTGCTTCAAGAGCTTCTAAGAGGCTGTGCTGGGCTGTG 1243  
 QY 901 gccaggggtgaaactcccgcgcaataatctattgcttccatgacgaggtatccgagat 960  
 Db 1244 GCGAGGGGTGAAGCTCCCGCGCAATATCATTTGCGCTCTAATCGGGTTATCCGAGAT 1303  
 QY 961 taacagatgagcatctaacacagacggagagtgatacttaagaaacctcaagaagcc 1020  
 Db 1304 TACACGAGATGGCATCTGACACGACCGGAGGCTACTAGTGAAGACCTCTCAAGGAGGCC 1363  
 QY 1021 tccctga 1026  
 Db 1364 TCTCTGA 1369

LOCUS AF013571 5849 bp DNA linear BCT 03-JAN-1998  
 DEFINITION Thermus sp. Ys45 plasmid pTsp45s: Tsp45I restriction endonuclease (tsp45IR), Tsp45I methyltransferase (tsp45IM), and replication protein (repT) genes, complete cds.  
 ACCESSION AF013571  
 VERSION AF013571.1 GI:2735918  
 KEYWORDS Thermus sp. Ys45.  
 SOURCE Thermus sp. Ys45.  
 ORGANISM Bacteria: Thermus/Deinococcus group; Thermus group: Thermus.  
 REFERENCE 1 (bases 1 to 5849)  
 AUTHORS Wayne J. and Xu, S.-Y.  
 TITLE Identification of a thermophilic plasmid origin and its cloning within a new Thermus-E. coli shuttle vector  
 JOURNAL Gene 195 (2), 321-328 (1997)  
 MEDLINE 97449309  
 REFERENCE 2 (bases 1 to 5849)  
 AUTHORS Wayne J., Holden M., and Xu, S.-Y.  
 TITLE The Tsp45I restriction-modification system is plasmid-borne within its thermophilic host  
 JOURNAL Gene 202 (1-2), 83-88 (1997)  
 MEDLINE 98087421  
 REFERENCE 3 (bases 1 to 5849)  
 AUTHORS Wayne J., Holden M., and Xu, S.-Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-JUL-1997) Research, New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA  
 FEATURES  
 source  
 1..5849  
 /organism="Thermus sp. Ys45"  
 /strain="Ys45"  
 /db\_xref="taxon:64227"  
 /plasmid="pTsp45s"  
 <602..>1612  
 /gene="tsp45IR"  
 602..606  
 /gene="tsp45IR"  
 623..1612  
 /gene="tsp45IR"  
 /note="R 'tsp45I'"  
 /codon\_start=1  
 /transl\_table=1  
 /product="Tsp45I restriction endonuclease"  
 /protein\_id="AAC46043.1"  
 /db\_xref="GI:2735919"  
 /translation="MAEMWNTQRSVELLEKGYLDLQYKGESSSRSPPEVEEK  
 LREAYKAYEGRODPSPEATKLYEAVLNARKVERSPFNHPLVLYYVSEKAEKANK  
 ALERALOEVASKHPETIRVLAKENORGVALLORKEPEINROIGPMKRYKREL  
 KGTEERLPQPTKRIIVYSPKSPPOAPLIAERAGITTYGSDPAKDAKENUG  
 LGEEAEIGTKGVDPEYVYVIRSPETHTLIEVAFQSDFGGNQNDLVAKASIRLDE  
 KRHIGIVVDGMPYVSKFRGMAGIGKETTIVSVLLPLDLIAELHYQGEALGL"  
 complement(1609..>2860)  
 /gene="tsp45IM"  
 complement(1609..2850)  
 /gene="tsp45IM"  
 /note="M 'tsp45I'"  
 /codon\_start=1  
 /transl\_table=1  
 /product="Tsp45I methyltransferase"  
 /protein\_id="AAC46044.1"  
 /db\_xref="GI:2735920"  
 /translation="MSRSYPGLTRKAPLAKSKTSESPPRIVYPGKRDKEKLLDPT  
 OLVRKRETLFLGNAFLPEIDIGYFISGENQVLRKMLRPGCAGAGVOIYTDPT  
 YGIGQOFIVGDDTDVAVYISOPKNOIGDITLDSPQVEYELREHLITLRELMDSG  
 LIRVPHIDEXGDFEVKILIDBEYRRNFVNHIAIASPNPNSRKAAGSGOKMILVYSK  
 TRDIYVNESASPTSEELAKLPEYVDENERYTNPLNAPGRTKGPTGRPMRGILPP  
 PGRHWRYRPEKRLDELDAOGLIVMSKGVPRKRYADRLKKGKIKLQWQFDPYPR  
 YPTKELNDMLKTLIVQTSNEGDIIVDPFAGSGTTLIASPLLRKRSIGDASNEAVKAF  
 TRVLEDFPRLQHKFEIVSAF"  
 2164..4500  
 /note="required for plasmid replication (putative)"  
 complement(2856..2860)  
 /gene="tsp45IM"

RESULT 3  
 AF013571/c

rep\_origin  
 RBS

gene	complement(3132..>4169)
gene	/gene="rept"
CDS	complement(3132..4157)
	/gene="rept"
	/function="required for plasmid replication"
	/note="Rept"
	/codon_start=1
	/transl_table=11
	/product="replication protein"
	/protein_id="AAC6045.1"
	/db_xref="GI:2735921"
	/translation="MKNEKTFEELXEALETHNDNTATRGSDGSEDFLATPPDD KGLNRLAKFTYOKELRALTALEKNDHEAFSLSSVAPP1PPAEPPVGVACVADGQ KLLELDIAKSTIPYGNRELMRKVGVTFVPEMLLNLNGTFOYTHARKVLEKK GLVATDIDHOTVNGERRAIGTTLMAVRBLRGKARLTLDIYIPWRILADMANGVLSFN WKAIVQDHGIRPTIDVIVYMAOCKRVMPTKTVAVDGLIIVLPEVRSKTPALITLIL AVYATLADLDRSRFRFYAGLLMNAVANGELPAQYLTFAVLMKRVIRDTDSHLRPAALIV KTLKEAS"
misc_feature	complement(3200..3208)
	/gene="rept"
	/note="putative Dnaa box required for plasmid replication"
misc_feature	complement(3335..3343)
	/gene="rept"
	/note="putative Dnaa box required for plasmid replication"
RBS	complement(4165..4169)
	/gene="rept"
-10_signal	complement(4194..4199)
	/note="putative"
-35_signal	complement(4218..4223)
	/note="putative"
-10_signal	complement(4253..4258)
	/note="putative"
-35_signal	complement(4276..4281)
	/note="putative"
BASE COUNT	1377 a 1501 c 1670 g 1301 t
ORIGIN	

	Query Match	Similarity	Score	ID#	DB	Length	Mismatches	Indels	Gaps
Db	Best Local Similarity	100.0%	Pred. No.	2,7e-174					
	Matches 1026;	Conservative	0;						
Oy	1	gtgaagaacgaaaaaacctccttccttgaaagagcttaagagctcttagaggaaaccacgac	60						
Db	4157	GTGAAGAACGAAAAAACCTTCTTTGAAAGCGTTTCAGAGGGCTTTAGAGAAACCACGAC	4098						
Oy	61	aacacgaatgcactagggagtcacagataaggggtcaaggagctctctcttctgcccacgac	120						
	4097	AACCCATGCCACTAAGGGGTCAGTAAGGGGTCAGAGAACTTCTTTGGCACCGAC	4038						
Oy	121	ccccccacagatggaggtgcgaaaatcgctctcggaaggctttacctaccacaaaagag	180						
Db	4037	CCCCCTCACAATGAGAGTGCCGAAAATCCCTCCGCAAGGGCTTTACTATTACCAAAAAAG	3978						
Oy	181	gcaattggagtctgtcttaaccggaagaacacatagagctcttcctctcctctgttggggcc	240						
Db	3977	GCACTTAGAGATGCTTTACCCGAGAAAGACATAAGAGCTTCTCTCTGTTGGGGCC	3918						
Oy	241	cccccatcacccaacagcttgaaaccccccgcttgggaatgatatgtccaagccgtccaggaaggg	300						
Db	3917	CCCCCTATACCAACAGCTGTAACCCCCGTTGGGAATGTGTCAAAGCCGTCCAGACGGG	3858						
Oy	301	cctcaagaagctctcttggaacacctccacaaggagatggccgctccacacatccctacgagcaac	360						
Db	3857	CCTCAGAAAGCTTCTGGAACTCTTCAGAGAAATTGCCCGCTCACCATCCCCTAGGGCAAC	3798						
Oy	361	cggagagctcttgaaagaaagtggggagcggtcgtctcctaaggtcccccvtggaatgttggcc	420						
Db	3797	CGGAGAGCTCTGAGAGAGAGTGCGGAGACGTCGTCCTTATGTCCTCCCCTGGAATGTTGGCC	3738						
Oy	421	ctcaaacctggggggtccaccggaagaaacggtccacagccttgtaaagaagtgctcttgaagaag	480						
Db	3737	CTCAACCTGGGGGTACCCCGGACACCGTCACAGCCTTGGAAGAAGTGCTTTSAGAAAAAG	3678						

QY	481	gacctcgttggccaccgagcgcctccttaccaaaacggttaacgaggagagccggcgccatcggc	540
Db	3677	GGCCTGGTGGCCACCGACGCTCTTCCACCAACCGTCAACCGGGAGGCCCGGGCCATCGGC	3618
QY	541	aacctctgggcccgtccgagctggaagccaggggaaagccagagctcaacctgagcgaactaac	600
Db	3617	ACCTTTTGGGCGCTCCGGCTGAGGGCCAGGGGAAGCCAGGGCTCACCCCTGGACGACTACATC	3558
QY	601	tacccttggaggaaacctcgccctctagacatggtccaaagcggtgtctctcttcaactggtc	660
Db	3557	TACCCCTGGAGAACCTCGCCCTAGACATGGCCCAACGGCGGTCTCTTCAACTGGGTC	3498
QY	661	aagcgcttaccagagacacagagatccgcgcccaacctgagcgtgtcgtcctctggtcctcag	720
Db	3497	AAGGCTTACCGAGACACAGCAATCCGCCCCACCCCTGGACGCGTGGTCTGTGGGCTCAG	3438
QY	721	gggaaaaaggtgatgtgccacaacacagaacgcttgacgttgacgttgacgttgacgttgac	780
Db	3437	GGGAAAAGGGGTGATGGCCCMACCAAGACCGTGGCCGTGACCTGGGGCTTCATCTGTGTC	3378
QY	781	ctcccgagaggtggagagcgtctcgaactcccgcccttataccacctcatgtgtacgtacat	840
Db	3377	CTCCCGAGGGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCCATGTGTACGTACATTT	3318
QY	841	gcccgtctccctagatgacgctgcttcaagagcgttctctatgacagcgttgcttggtgctgtg	900
Db	3317	GCCGATCTCTTAGATGACCGGTCTCAAGAGCGTTCTATGACAGCGCTTGCCTGTGGCGTGTG	3258
QY	901	gcccagaggttgactcccccgcgccaatatctatttgcgtccctcaatgagcgttgatccagat	960
Db	3257	GCCAGGGGTGAACCTCCCGCCCAATATATTATTGGCGCTCTAATAGCGGGTTATCCGAGAT	3198
QY	961	tacacggaatgagccatctgacacgacacgggagcgtaactagtgaagaacctcaaggagggcc	1020
Db	3197	TACACGGATGGCCATCTGACACGACCGGGAGCGTACCTAGTGAAGAAGCCCTCAAGGAGGCC	3138
QY	1021	tccctga 1026	
Db	3137	TCCTGA 3132	
RESULT 4			
LOCUS	ARI39426/c	5849 bp	DNA linear PAT 16-JUN-2001
DEFINITION	Sequence 6 from patent US 6207377.		
ACCESSION	ARI39426		
VERSION	ARI39426.1	GI:14481922	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 5849)		
AUTHORS	Wayne, J. and Xu, S.-Y.		
TITLE	Method for construction of thermus-E. coli shuttle vectors and identification of two thermus plasmid replication origins		
JOURNAL	Patent: US 6207377-A 6 27-MAR-2001;		
FEATURES	Location/Qualifiers		
source	1..5849		
BASE COUNT	1377 a 1501 c 1670 g 1301 t		
ORIGIN	/organism="unknown"		

Query Match	100.0%	Score 1026	DB 6	Length 5849
Best Local Similarity	100.0%	Pred. No. 2.7e-174		
Matches 1026	Conservative	0	Mismatches 0	Indels 0
QY 1	gtgaagaacgaaaaaacctcttgaagagccttaagagccttagaggaaccacgac	60		
Db 4157	GTGAAGAACGAAAAAACCTCTTTGAAGACCTTTACGAGCCTTTAAGAGCAACCCACGAC	4098		
QY 61	aacacgcagtcacacttagggggctcagataggggctcagagagactcttcttgccacgac	120		

```

DB 4097 AACACGAGTCCACTAGAGGGGGTTCAGTAGGGGGTTCAGAGACTCTTCTGGCCACGAC 4038
QY 121 cccctccagatgaggtgcggaataatcgctcgcgaaggtcttacataccaaagaag 180
DB 4037 CCCCTCCAGATGAGAGTCCGGAATCCCTCGCGAAGGGCTTTCATACCAAAAAAGAG 3978
QY 181 gaacttagattgcttaccgcgaagaagacattgaagcttctctctctgttgggccc 240
DB 3977 GCACCTTAGATTCCTTACCCTCGAAGAACATGAGCGCTTCTCTTCTCTGTGGGGCC 3918
QY 241 cccctatacccaacagatgaacccccgttgggaatgatagtcaacgcgtccagagacgg 300
DB 3917 CCCCCATACACACACTGTAACCCCCCGTGGGAATGTAATGTAACCCGTCGAGACGG 3858
QY 301 cctcagaagcttctggaactcctccagagatgtccgcgtccacacatccctcagggcaac 360
DB 3857 CCTCAGAACCTTCTGGAACCTCTCCAGAGATTGCCCGCTCCACATCCCTACGGCAAC 3798
QY 361 cggagagctctgagagaggttgggagagctgcttcaatgctccctcagagatgttggcc 420
DB 3797 CGGAGACTCTGAGAGAGGTGGGGAGGCTGCTCTCATGCTCCCTCGAGATGTTGGCC 3738
QY 421 ctcaacctgggggtlcaaccggcagaccgtlccacgcttggagaaggctccttgaagaaag 480
DB 3737 CTCACACTGGGGGTCAACCCGGCAGACCGTCACGCTGGAAGAGTCTTGAAGAAAG 3678
QY 481 ggcctgtgtgcaacgcagctccttcaacaaacgctaaaggggagagcgccgacatcgcc 540
DB 3677 GGCCTGTGTGCCACCGACGCTCTTCAACAAACGCTCAACCGGAGGCGCCGGGCATCGGC 3618
QY 541 acccttgggcgcgtccgcgtgagggcgaaggaagcagagctacccctggagagctacatc 600
DB 3617 ACCCTTGGGCGCTCCGCTGAGAGGCCAGGGAAGCCAGGCTCACCTGGAGCATACATC 3558
QY 601 taacctggagagacctcgccttagacatlggccaagcgctgtcctccttcaactgggttc 660
DB 3557 TACCCCGAGGAAACCTCGCCTTAGACATGAGCCAAACGGCGTGTCTCTCAACTGAGGTC 3498
QY 661 aaggtctaccagagcaacggaatccgccaccctcgtgagctgtgtctccttgggtcag 720
DB 3497 AAGGCTTACCAAGACACGGAATCCGCCACCTGACGCTGTGGTCTCTTGGGCTCAG 3438
QY 721 ggggaaaggttgatgcccaacacaaagacgttgacgttgacgtgggcccacatcgtgtc 780
DB 3437 GGGAAAGGGTGAATGCCCAACACCAACGCTGAGCTGAGCTGGGCTCATCTCTGTC 3378
QY 781 ctccccaagttgagcgtlccaaactccgcgccttaccctcctacatctgtacgtacat 840
DB 3377 CTCGCCGAGGTGAGCGCTTCCAAACCTCGGCGCTTATACACCTCATTTGCTAGCTACAT 3318
QY 841 gcgagatcctagatgacgctgcttcaagacgcttctacagagctgtcgttgggtcgtg 900
DB 3317 GCCGATCTCTAGATGACCGCTCGTTCAAGAGCTTCTATGCAAGCTGTGCTGGGCTGTG 3258
QY 901 gccaggggtgaacccccgcgcaatatctatcttgcgtcccaatgcggtatccagat 960
DB 3257 GCCAGGGGTGAATCCCCCGCAATATCTATTTCCTCTTAATGCGGGTATCCGAGAT 3198
QY 961 taacagatgagcactcgaacagacggagcgctacatgaaagacccctcaagagagcc 1020
DB 3197 TACACGAGTGGCCATCTGACACGACGAGCGGAGCGTACTAGTGAAGACCTCAAGAGGCC 3138
QY 1021 tctctga 1026
DB 3137 TCTCTGA 3132

```

```

RESULT 5
SC2E1
LOCUS SC2E1 38962 bp DNA linear BCT 04-JUN-1998
DEFINITION Streptomyces coelicolor cosmid 2E1.
ACCESSION AL023797

```

## VERSION

AL023797.1 GI:3191976  
30S ribosomal protein S16; 30S ribosomal protein S2; 50S ribosomal protein L19; cell division protein FtsH homolog; elongation factor

## KEYWORDS

fts; fth; ftsH; glnB; glnD; nitrogen regulatory protein PII; prophase; protein pili uridylyltransferase; pyrH; RNA polymerase sigma factor whiG; rplS; rpsB; rpsP; signal peptidase I; signal recognition particle protein; sip1; sip2; sip3; sip4; trmD; tRNA (guanine-N1)-methyltransferase; tsf; uridylylate kinase; whiG.

## SOURCE

Streptomyces coelicolor A3(2).

## ORGANISM

Streptomyces coelicolor A3(2).

## REFERENCE

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomyces; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

## AUTHORS

1. (bases 1 to 38962)  
Murphy, L. and Harris, D.

## JOURNAL

2. (bases 1 to 38962)  
Parkhill, J., Barrell, B.G. and Rastandream, M.A.

## REFERENCE

Submitted (03-JUN-1998) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrellesanger.ac.uk Cosmids supplied by Prof. David A. Hopwood. [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

## AUTHORS

3. (bases 1 to 38962)  
Redenbach, M., Kleiser, H.M., Denapalte, D., Eichner, A., Cullum, J., Kinash, H. and Hopwood, D.A.

## TITLE

A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome

## JOURNAL

Mol. Microbiol. 21 (1), 77-96 (1996)

## MEDLINE

97000351

## COMMENT

Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.

## REFERENCE

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>) CDS are numbered using the following system eg SC2B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

## AUTHORS

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

## TITLE

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or atc) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

## JOURNAL

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 2E1 lies between 7A1 and 6A9 on the AseI-B genomic restriction fragment.

## MEDLINE

location/Qualifiers

## COMMENT

1. 38962  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid 2E1"  
1. 166  
/gene="glnB"  
1. 166  
/gene="glnB"  
/note="SC2E1.01, glnB, nitrogen regulatory protein pili,

## FEATURES

FEATURES

## SOURCE

1. 38962  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid 2E1"  
1. 166  
/gene="glnB"  
1. 166  
/gene="glnB"  
/note="SC2E1.01, glnB, nitrogen regulatory protein pili,

## CDS

1. 166  
/gene="glnB"  
/note="SC2E1.01, glnB, nitrogen regulatory protein pili,

## gene

1. 166  
/gene="glnB"  
/note="SC2E1.01, glnB, nitrogen regulatory protein pili,

## CDS

1. 166  
/gene="glnB"  
/note="SC2E1.01, glnB, nitrogen regulatory protein pili,

```

partial CDS, len: >54 aa; highly similar to many e.g.
GLNB_AAOBR (112 aa), fasta scores: opt: 198 z-score: 381.9
E(): 4.9e-14, 50.0% identity in 54 aa overlap"
/codon_start=2
/transl_table=11
/product="nitrogen regulatory protein PII"
/protein_id="CAA19376.1"
/db_xref="GI:3191977"
/translation="SPTREMBL:069872"
/feature="IRIEVLVEDDDAQLIDVYKARTGKIGDKWAVYDTAVRY
KTCGRGPDAI"
231..2738
/gene="glnD"
231..2738
/gene="glnD"
/note="SC2E1_02, probable glnD, protein PII
uridylyltransferase, len: 835 aa; similar to many e.g.
GLND_ECOLI (890 aa), fasta scores: opt: 403 z-score: 540.4
E(): 7.3e-23, 26.2% identity in 851 aa overlap. Contains 7
degenerate repeats of the sequence PSS"
/codon_start=1
/transl_table=11
/product="putative protein PII uridylyltransferase"
/protein_id="CAA19377.1"
/db_xref="GI:3191978"
/translation="MTDEADSGPGGYAARLRLLTEGASGPPRRARALAEIDGMIA
GLFCATEEHTGISLVAGVGREGELSPSDILLLHGDROKAVAAIADRLPYPM
DLGIDHDSVTRPQOARKTAGODKYLGLDLDHRLHAGDGLTRSLTAYLADBRNOA
PKRLPELRDCAERAROGELQFLLEPDLKARGLKLDATRLARVAASWLDAREGL
AEARRLLADVDALHATGRATDRLALQEDQVAELGLDADALRLQVYEAARVLSY
AGDWTREVRVLSRSRPRRLAMNMGKGVKVAERSEVQDEAVLARIYA
RPERDPALPLRAAAAGLPLSRHVRNLAATARLPPTMPPEAREOVLTLGSR
PTVQVWALAEGLVTRLLPDWERVRCRPNRNVAHWTVDRLHLETRVRAAGFRRVH
RPLLLALHLDIGKMGPDHVSAGETIARDVAARIGPDADTAVATLTVRHLLIY
ETATRRDLDPATYRAVAOVAGTEHTLHLEALTEADALTCGAAASMGSLVADIV
KYSVCVAGSPQEPABEAPTAEDERLAVAFRTGGVPLRLAQTEPPADSPAPSSP
SSPSPSPSSPSSADQPELVGELLTAVPDQAVLPAVAGVLAHMLVYRTAL
RSVPLPDGVGSVLLDMWVAAGVSLPQAARLRADRLDGLTDAARLAERDAH
PRRGVPPRPRTVAPAAASRLATVLEVRQADAGLFLRGLRALAEAGVRSAAHVS
LGNADVAEFTVTEGEGTPLPDDEASVARGLESIRT"
2091..2153
/note="7 degenerate 9bp repeats"
2719..2723
/gene="glnD"
/note="possible RBS upstream of ffn"
2735..4387
/gene="ffn"
2735..4387
/gene="ffn"
/note="SC2E1_03, ffn, signal recognition particle protein,
len: 550 aa; highly similar to may e.g. SR54_ECOLI (453
aa), fasta scores: opt: 1465 z-score: 1282.8 E(): 0, 50.0%
identity in 458 aa overlap. Contains PS00017
ATP/GTP-binding site motif A (P-loop), PS00300 SRP54-type
protein GTP-binding domain signature and Pfam match to
entry SRP54 PF00448; SRP54-type proteins, score 66.57.
Similarities suggest possible start site at aa 35"
/codon_start=1
/transl_table=11
/product="signal recognition particle protein"
/protein_id="CAA19378.1"
/db_xref="GI:3191979"
/db_xref="SPTREMBL:069874"
/translation="MTVRSVORAGTCLCPGRITPMTTTHNADARTYAIVEDLSDRLS
ATPESLRGKGRLEADIDATAREIRIALLEADVAPVAFIKKVERSLGAWSKAL
NPAQOVLYKIVNEELVIGGETRRLRPAOPPVIMALGLOGKRTTLAGKIGHWKSL
OGSHPLVACDLDORNAVNOLSVABRACAYAPRPGNGVGPVYVAADSIETFAK
VHDLVYVTRAGRTGIDQELMOQADTRDAVSPDEILFVDDAMTGDAVNTAEFRDGV
GPDVAVLSKIDGARGAALSLASVTKPIMFNSGKELDDMFADPHDRMAEFLDMG
DLTLLEQAEKTESQEAERMAKSLASKSGODETDLFLAQMDFWRKMSISKLGLML
PGMGOKDDOINNIDERDVRTAALITKSMTPGEQDEPTIINGSRARITAGSGVEAV
KNLVERFFEARKKMSMAOGGGMWPMGMWPMGMGGRQKKKKKAKKGRORSNPKRK

```

```

misc-feature
3155..3178
/gene="ffn"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
3566..3697
/gene="ffn"
/note="Pfam match to entry SRP54 PF00448, SRP54-type
proteins, score 66.57"
misc-feature
3647..3668
/gene="ffn"
/note="PS00300 SRP54-type proteins GTP-binding domain
signature"
4405..4453
/note="hairpin loop with 22/23 bp stem"
4504..6450
/gene="f1sh"
4504..6450
/gene="f1sh"
4504..6450
/gene="f1sh"
/note="SC2E1_04, f1sh, cell division protein f1sh homolog
(zinc metalloprotease, integral membrane protein), len:
648 aa; similar to many e.g. FTSH_LACLA (695 aa), fasta
scores: opt: 1489 z-score: 1195.5 E(): 0, 41.7% identity
in 640 aa overlap. Contains PS00017 ATP/GTP-binding site
motif A (P-loop), PS00674 AAA-protein family signature and
Pfam matches to entry AAA PF00004, ATPases associated with
various cellular activities (AAA) and to entry zn-protease
PF00099, zinc-binding metalloprotease domain, score 22.07,
score 357.96. Proline-rich N-terminus is not present in
other f1sh homologs"
/codon_start=1
/transl_table=11
/product="cell division protein f1sh homolog"
/protein_id="CAA19379.1"
/db_xref="GI:3191980"
/db_xref="SPTREMBL:069875"
/translation="MTNPSPRRKAPRPPMTTGTGTPDEPPKPPRGRRMGNNLILA
ALIVYLIANLVLSFNEGDEPTLSYTESKOVDEGVSKIYAGDAIQGLKARNP-
EGGTYTKFTNRTVTPADQIMLDKRNRTVAPVQVRSFLANLILALMLIV
VLMIFARRRGALGGAGGMLKMDKPRKVELEAKPRTPTADVGDIVEBELSDY
VDLKNDAVARRGAKRGVLLTGPGCTGTLTALAAGVAGVPPFSAASPTIMI
VCGASVRELPFAEAKKRVASITFIDEIDTIGARRGSGSTGSHDEREOTLNTIEM
DGRSGSGVIVATNRADILDAALTRPGFDFVAVSPDRGRALILEIHTREIP
APDIDIAQVARTTPGMGALANLANEALAVKROERTVQNLSEALEKVOUGAER
PLVMPBEERRRTAVHESGALLGMLQPGADPVKRTIIVPRGALVTLSIPDADKAY
TTEYLGRITIGALGMAAEHVYGMITTSSESLDEVNTNARGMVARMGSESRVGLS
ALPGDQOAGYLAAPOTLDAIDGEMRRVVDSCYEAVKRLRDRQDALAEASLAS
ETLDEADAVYRIAGTITRLKDDPRA"
5215..5781
/gene="f1sh"
/note="Pfam match to entry AAA PF00004, ATPases associated
with various cellular activities (AAA), score 357.96"
misc-feature
5215..5781
/gene="f1sh"
/note="Pfam match to entry AAA PF00004, ATPases associated
with various cellular activities (AAA), score 357.96"
Query Match
5.8% Score 59.8; DB 1; Length 38962;
Best Local Similarity 46.3%; Pred. No. 0.22;
Matches 196; Conservative 0; Mismatches 227; Indels 0; Gaps 0;
QY 374 gaaagtggaagcgtgcttcacgtgccccctgagagatggtgcccccaacctgaggg 433
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 1423 GGGAGGCGCGTGAAGCGGAGGCGGTGTCACCGCGTGTGACCGGACATGGGAGCGGTC 1482
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 434 tcaacggagcagccgttcacgacctggaagaagtccttggaagaagagcctgtgcca 493
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 1483 GCTGCCCGCCCGCAGCGCACCGCTGACCTGTGACCGTGGACCGCGACCTGATCAGA 1542
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 494 ccgaagctctcaacaaacggtcaacgggagagcgcgagcaccgacaccttgagcg 553
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 1543 CCGCGCTCCCGCGCGCGGCTTACCGCGGCTGCACCGCCCGACACTGCTCTCTGATCG 1602
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 554 tccgctgagcagcaggaagcagcagccttgacacctacatcctacacctgagga 613
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 1603 CCGGCGTGTGTCACGACATGCGCAAGGCGTGGCGCGCGACACCTCGTGGCGGAGGA 1662
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 614 acctgcacctagacatgagcagcgctgctctccttcaactgagtgcaagcgctaccagg 673
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 1663 CCATGCCCGGACGCTCGCGCGCCGATCGCTTGACGCGCGGACACAGCGCGCTCG 1722  
 Oy 674 accacggaatccgccaccctggaactgctgctctgtggtcctcagggaaaggtga 733  
 Db 1723 CCACCTCTGTACGACACCTCTCTGCTGTGAGACGCCACCGCGCGGACCTGGAG 1782  
 Oy 734 tggccaaacccagacgctgctgacctgacctgctcctcctcccccaggtg 793  
 Db 1783 ATCCGCGCCAGCTGCGCGCGGCGGACGCGGCGGACGACACGCTGAGCTGC 1842  
 Oy 794 agc 796  
 Db 1843 TGC 1845

RESULT 6  
 SCC42/c 13347 bp DNA linear BCT 27-MAY-2000  
 LOCUS Streptomyces coelicolor cosmid C42.  
 DEFINITION AL356992.1 GI:8247035  
 ACCESSION ABC-transporter ATP-binding protein; ABC-transporter integral  
 VERSION membrane protein; phosphate binding protein; regulatory protein;  
 KEYWORDS secreted beta-galactosidase; secreted protein.  
 SOURCE Streptomyces coelicolor A3(2).  
 ORGANISM Streptomyces coelicolor A3(2).  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.  
 1 (bases 1 to 13347)  
 Redenbach, M., Kleser, H.M., Denaplatte, D., Elchner, A., Cullum, J.,  
 Kinashi, H. and Hopwood, D.A.  
 A set of ordered cosmids and a detailed genetic and physical map  
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
 Mol. Microbiol. 21 (1), 77-96 (1996)  
 2 (bases 1 to 13347)  
 Brown, S.P. and Harris, D.  
 Unpublished  
 3 (bases 1 to 13347)  
 Cerdano, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.  
 Direct Submission  
 Submitted (26-MAY-2000) Streptomyces coelicolor sequencing project,  
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
 Colney, Norwich, Norfolk NR4 7UH, UK

## COMMENT

Notes:  
 Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BSRG and Beowulf Genomics  
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL: <http://www.sanger.ac.uk/Projects/S-coelicolor/>)  
 CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
 The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
 Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most

upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
 Cosmid C42 overlaps with cosmid C24 on the AseI-C.  
 location/Qualifiers

## FEATURES

## source

## misc-feature

## CDS

1..13347  
 /organism="Streptomyces coelicolor A3(2)."  
 /strain="A3(2)."  
 /db\_xref="taxon:100226"  
 /clone="cosmid C42"  
 1..99  
 /note="nominal overlap with Streptomyces coelicolor cosmid Str82"  
 complement(1..486)  
 /gene="SCC42.01c", possible ABC-transporter integral membrane protein (fragment), len: >161 aa; similar to TR:053545 (EMBL:AL022022) Mycobacterium tuberculosis hypothetical 29.6 kD protein MY023.07c, 280 aa; fasta scores: opt: 565 z-score: 630.0 E(): 1.3e-27; 54.58 identity in 154 aa overlap. Contains possible hydrophobic membrane spanning regions"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative ABC-transporter integral membrane protein (fragment)"  
 /protein\_id="CAB92894.1"  
 /db\_xref="GI:8247036"  
 /translation="MASPLVWLDNRSGDQLFLYRALLWVPTLRRLKEVORLAEVA FESGGLGVGTIGVMIAMLTLETVGVGIGVAAADIGTSAFGEFVAFNTRTAP LVAGIALATVAGFTAQDAGMRINEEDALEMGIRSMPLYLTTRIIAGVAILPLY A1"  
 complement(1..486)  
 /gene="SCC42.01c"  
 complement(491..494)  
 complement(493..1257)  
 /gene="SCC42.02c"  
 complement(493..1257)  
 /gene="SCC42.02c"  
 /note="SCC42.02c, possible ABC-transporter integral membrane protein, len: 254 aa; similar to TR:053546 (EMBL:AL022022) Mycobacterium tuberculosis hypothetical 26.6 kD protein MY023.08c, 254 aa; fasta scores: opt: 881 z-score: 1004.0 E(): 0; 50.88 identity in 242 aa overlap. Contains possible hydrophobic membrane spanning regions"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative ABC-transporter integral membrane protein"  
 /protein\_id="CAB92895.1"  
 /db\_xref="GI:8247037"  
 /translation="MTTGLRGTGRFLAARVAVRPPRPPREFVRFVAVST IIPALAVISIPFGAVINLVGSLTEIDGAGSFTFGASVLAIVQASPLVAILIACAG SAICADIGSRRIRELDIAVMGVSPVQSLVPRVLAAGVALLGLVSVGIIAGT FGNVIMOGCTPGAVLISFSAALMDPLVSEIKALVFGIAGIIVAAVRLNDRGPKG VGDVAVNOGVITFILLFVNMWMTAVYLIQIVPPKG"  
 complement(1254..2285)  
 /gene="SCC42.03c"  
 complement(1254..2285)  
 /gene="SCC42.03c"  
 /note="SCC42.03c, possible ABC-transporter ATP-binding protein, len: 343 aa; similar to SW:YRBF\_HAEIN (EMBL:U32788) Haemophilus influenzae hypothetical ABC-transporter ATP-binding protein H11087, 264 aa; fasta scores: opt: 680 z-score: 686.6 E(): 8.9e-31; 41.4% identity in 244 aa overlap. Contains Pfam match to entry PF00005 ABC\_tran, ABC transporter and matches to Prosite entries PS00017 ATP/GTP-binding site motif A (P-loop) and PS00211 ABC transporters family signature"  
 /codon\_start=1  
 /transl\_table=11

```

/product="putative ABC-transporter ATP-binding protein"
/db_xref="GI:8247038"
/translation="WASAGDMGIEVVEVGLTKSFKOSIMRDVSLTLPAGEVSMGL
PSGKIVFELKSLGILKPEKGRVLIINCIVANSPERDIYERKLFGLMFGDGLFSG
MSLEPNVAFPLERHEKRESEIRIYMRIVEMVNLGAEKLPKPGISGMKRGGLAR
ALVIDPQIILCDEPDSGLDPVTAIISOLLIDLANQDATTMLIYTHNIDIAIYDNN
GMLFLRQVTEGPRVGLTSDPEVAOPIGGRRCRPIGMSSEKDAITIAADAPAPA
PAAPRVYPOLEPSRGLPRRAVARRRKRVGMIDTLLPAAKSAIRDTYAROSEATLL
PMPAAGSAA"
RBS
complement(1264..1267)
/gene="SCC42.03c"
complement(1620..2177)
/misc-feature
/gene="SCC42.03c"
note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 152.90, E-value 5.4e-42"
complement(1800..1844)
/gene="SCC42.03c"
note="PS00211 ABC transporters family signature"
complement(2133..2156)
/misc-feature
/gene="SCC42.03c"
note="PS00017 ATP/GTP-binding site motif A (P-loop)"
2904..2909
RBS
2918..3350
/gene="SCC42.04"
2918..3350
CDS
/gene="SCC42.04"
note="SCC42.04"
Contains possible N-terminal region signal peptide
sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAB92897.1"
/db_xref="GI:8247039"
/translation="MKKNMRSAIYVCAASALGLAAPPSSAVPSTVWTVNPSPAEK
AVANGNIVLRTIPCTCTTAAGTMAASAGNPGDVADITTMNGASGSPCTSYLGNV
ITTSVTWADVVDVNSGTGKGVNANVAGACKFTYMGKAGTITNSTGYLA
INSTAGELAVSNPVCALVTTATKPTFKGNVAVTASTGAIPTVGSNP"
3679..3682
RBS
3687..5288
/gene="SCC42.05"
3687..5288
CDS
/gene="SCC42.05"
note="SCC42.05"
Contains possible N-terminal region signal peptide
sequence. High content in alanine amino acid residues"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAB92898.1"
/db_xref="GI:8247040"
/translation="MKATPAAPRPSRARATTSIAFVIALVPTTASAGTOEVE
AELPYCTLPSGLPATVRVSAEPERAGADEFTPDVTTVELPAEAVADITADBA
AERAAATSLAVGAONATTAATVWRGSAEPVALPGSRPLTVRGVPSVAGSDGL
TFSAGALATIDLAGADPRATDPSGLTVDCTLDEAGGLLAVPGVGTGQAPSSGP
SSSGPAGSSGAPDDGRQDGPGRPRRSESPKYLENPGAAARDVDVPCRYDEO
HPPLVDLNAVTVGAVANWKKGAAYLPPSCVILEGLPVPGPDEYLLFDLSTAN
FHYRERKQTPFEATFLSPDFAVKATVLEQTKMRKIDSRKILSDFTTIDTVAN
APLVLVLDLEVNGTDLVSGSECTRTSLSEDPDNPEDHLLVLYGREOLIGLPA
TGYLLSGALGSEATIPATFGCGSDGDDRLTLTASVSGNVIQVQCQTCAIAP
VFPSPENEGCCTEDQPELPAER"
5323..6264
gene
5323..6264
CDS
5323..6264
/gene="SCC42.06"
5323..6264
/gene="SCC42.06"
note="SCC42.06"
Contains possible N-terminal region signal peptide
sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"

```

```

Query Match      5.5%; Score 56.8; DB 1; Length 13347;
Best Local Similarity 45.6%; Pred. No. 0.89;
Matches 199; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

RESULT 7
LOCUS      SC2D46
DEFINITION Streptomyces coelicolor cosmid 2D46.
ACCESSION  AL391406
VERSION     AL391406.1 GI:9857143
KEYWORDS    DNA-binding; GntR-family transcriptional regulator; hydrolase;
            integral; integrase; MarR-family; NRP/P60-family secreted; pgm;
            phosphoglycerate mutase; TelR-family; transferase; transport.
SOURCE      Streptomyces coelicolor A3(2).
ORGANISM    Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Streptomyces coelicolor A3(2).
REFERENCE   1 (bases 1 to 35576)
AUTHORS    Redenbach,M., Kleaer,H.M., Denapate,D., Eichner,A., Cullum,J.,
            Kinasli,H. and Hopwood,D.A.
TITLE      A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL     Mol. Microbiol. 21 (1), 77-96 (1996)
REFERENCE   2 (bases 1 to 35576)
AUTHORS    Brown,S.P. and Harris,D.
TITLE      Unpublished
JOURNAL     3 (bases 1 to 35576)
AUTHORS    Bentley,S.D., Parkhill,J., Barrell,B.G. and Randsdram,M.A.
TITLE      Direct Submission
JOURNAL     Submitted (03-AUG-2000) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at the Sanger Centre are

```



available on the World Wide Web.  
(URL: [http://www.sanger.ac.uk/Projects/S\\_coeilicolar/](http://www.sanger.ac.uk/Projects/S_coeilicolar/))  
CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid 2D46 lies between and overlaps cosmids D66 and D8 on the AseI-D genomic restriction fragment.

## FEATURES

## source

1. 35576

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db\_xref="taxon:100226"

/clone="cosmid 2D46"

complement(1..135)

/gene="2SCD46.01c"

## repeat\_unit

1..1200

## CDS

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

/note="2SCD46.01c, unknown, partial CDS, len: >45aa; strongly similar to neighbouring CDS 2SCD46.03c fasta scores: opt: 169, z-score: 236.7, E(): 1.4e-07, 51.1%

identity in 45 aa overlap. Hydrophobic."

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="CAC04023.1"

/db\_xref="GI:9857144"

/translation="MPLAIVAAALAGPEQLVQWKYGPWGIAFVALSIGIKANTM I"

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

gene

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## gene

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar to neighbouring CDS 2SCD46.04 fasta scores: opt: 1074, z-score: 984.2, E(): 0, 57.6% identity in 302 aa overlap. Contains Pfam match to entry PF00392 gntR, Bacterial regulatory proteins, gntR family and helix-turn-helix motif (Score 1967 (+5.89 SD))

## RBS

## CDS

misc-feature

1..98

/note="Nominal overlap with Streptomyces coelicolor cosmid D66 (EMBL:AI358692)."

complement(142..147)

289..1200

/gene="2SCD46.02"

289..1200

/note="2SCD46.02, possible GntR-family transcriptional regulator, len: 303aa; similar to many eg. TR:Q9RIQ2 (EMBL:A0243257) Kora protein from plasmid pSM1 of Streptomyces natalensis (245 aa) fasta scores: opt: 175, z-score: 204.2, E(): 0.00069, 28.2% identity in 238 aa overlap. Also strongly similar

/gene="2SCD46.05"  
 /note="2SCD46.05, possible hydrolase, len: 273aa; weakly similar to many proposed phosphatases eg. TR:CA84916 (EMBL:AL162756) putative phosphatase from Neisseria meningitidis (235 aa) fasta scores; opt: 195, z-score: 238.2, E(): 8.8e-06, 29.5% identity in 237 aa overlap. Contains pfam match to entry PF00702 Hydrolase, holoacid dehalogenase-like hydrolase. Note possible alternative downstream translational start codons."  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative hydrolase"  
 /protein\_id="CAC04027.1"  
 /db\_xref="GI:9857148"  
 /translation="MVRRLPLGNHHHDTLLVTSIDQEPVSERTGRDSEALNDL VGVRRVLDPEFGPICRLFAGYSADRVAGELVMELELKEILLTOEDVHPDVL AVVDRRRSDVLAEEFERLRELRVAMPVAYADALRTWSALGVAVTNS PRVSEYLERDRLGCFAPHLIGRTGDPHLKPDPHCLNLSAMGAAPRALVGD S ASDYTAARRAGVPEPLTGHNBRKTKLKQAGATVYDLSLEPVLRLMENTPAPA"  
 2504. .3076

misc\_feature 5.3%; Score 54.6; DB 1; Length 35576;

Query Match Best Local Similarity 46.6%; Pred. No. 1.9; Mismatches 199; Indels 0; Gaps 0;

QY 416 tggccctcaacctgggggtaccgcgagacccgtccagcgttggaagaagtcttga 475  
 16308 TGAACGTCTACATCGTGGAGCTGGCGACGCCCTCGCCGATCAACATCGAGTCA 16367  
 QY 476 aaaaagggcctgtgtgcccgcgcgcgtccctaccacacgcgtcaagcgagcgagcga 535  
 16368 TCTTCACCCGGGACGACACCGCCGCTCCGCCGCTGCAACTGGCCCCGGCGTCC 16427  
 QY 536 tcggcaccccttggcgctgcgctgagggccaggaagcagcgtccacccctgagcact 595  
 16428 TCGTCGGGACGTCACGCGGCGCGCTACGAGGCTGGCGCAAGAGAGAGCTGCCGCC 16487  
 QY 596 acatcacccctggagagaaacctgcgcctatgaatggcgaagcgctgtctcttcaact 655  
 16488 AGCTTCGCCCTTACCCACCGGCGTGAATCAAGCCTGGGCGGCTACCCCGGCGCACT 16547  
 QY 656 gggtaagggcctacacgagacacgcgcgcacacccctggagcgtgctgtctcttga 715  
 16548 ACACCTCGTGCACCTCCACACTACTGCTCTCGGCGCACTCGGCTGCGCGCCACAC 16607  
 QY 716 ctcaagggaaaggtgatgccaacacacgaacgcgtggcgttgacccctgacatcc 775  
 16608 GCTGGGCGCGCCCTGTGTGACACGCGCATGCACACATGGCCCAAGGTCAGAACGCCAC 16667  
 QY 776 tggctctcccgga 788  
 16668 TGGCCGACGGCGA 16680

RESULT 8  
 SCDS 41622 bp DNA linear BCT 21-SHP-1999  
 LOCUS Streptomyces coelicolor cosmid D25.  
 DEFINITION AL118514  
 ACCESSION AL118514 GI:5918466  
 VERSION ABC transport protein, ATP-binding subunit;  
 KEYWORDS ABC transport protein, ATP-dependent subunit;  
 ATP-dependent RNA helicase; DNA polymerase III subunit gamma;  
 DNA-binding protein; integral membrane protein; lipoprotein;  
 LSR2-like protein; luxR family; phosphoribosyl formylglycinamidine  
 synthase II; phosphoribosylamine-lysine ligase;  
 phosphoribosylaminimidazole-succinocarboxamide synthase;  
 phosphoribosylformyl glycinamidine synthase I; secreted peptidase;  
 sensor kinase; transposase; tRNA; two-component transcriptional  
 regulator; valine dehydrogenase.  
 SOURCE Streptomyces coelicolor A3(2).  
 ORGANISM Streptomyces coelicolor A3(2)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 REFERENCE  
 TITLE  
 JOURNAL

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 1 (bases 1 to 41622)  
 Redenbach, M., Kiser, H. M., Denapate, D., Elchner, A., Cullum, J.,  
 Kinashi, H. and Hopwood, D. A.  
 A set of ordered cosmids and a detailed genetic and physical map  
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
 Mol. Microbiol. 21 (1), 77-96 (1996)  
 97000351  
 2 (bases 1 to 41622)  
 Brown, S. P. and Harris, D.  
 Unpublished  
 3 (bases 1 to 41622)  
 Thomson, N. R., Parkhill, J., Barrell, B. G. and Raftery, M. A.  
 Direct Submission  
 Submitted (20-SEP-1999) Streptomyces coelicolor sequencing project,  
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
 Colney, Norwich, Norfolk NR4 7UH, UK  
 Notes:  
 Streptomyces coelicolor sequencing at The Sanger Centre is funded  
 by the BBSRC and Beowulf Genomics  
 Details of S. coelicolor sequencing at the Sanger Centre are  
 available on the World Wide Web.  
 (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)  
 CDS are numbered using the following system eg SC787.01c. SC (S.  
 coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary  
 strand).  
 The more significant matches with motifs in the PROSITE database  
 are also included but some of these may be fortuitous.  
 The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for  
 CDS which show significant similarity to other CDS in the database.  
 The position of possible ribosome binding site sequences are given  
 where these have been used to deduce the initiation codon.  
 Gene prediction is based on positional base preference in codons  
 using a specially developed Hidden Markov Model (Krogh et al.,  
 Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot  
 program of Bibb et al., Gene 30:157-66(1984) as implemented at  
<http://www.nih.go.jp/>  
 Jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the  
 correct initiation codon. Where possible we choose an initiation  
 codon (atg, gtg, tgg or (att)) which is preceded by an upstream  
 ribosome binding site sequence (optimally 5-13bp before the  
 initiation codon). If this cannot be identified we choose the most  
 upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the  
 sequenced clone. It may be shorter because we only sequence  
 overlapping sections once, or longer, because we arrange for a  
 small overlap between neighbouring submissions.  
 Cosmid D25 Overlaps with cosmid D17 on the AseI-D genomic  
 restriction fragment.

FEATURES  
 source

gene  
 CDS

1. 41622  
 /organism="Streptomyces coelicolor A3(2)"  
 /strain="A3(2)"  
 /db\_xref="taxon:100226"  
 /clone="cosmid D25"  
 /complement(1..451)  
 /gene="SCD25.01c"  
 /complement(1..451)  
 /gene="SCD25.01c"  
 /note="SCD25.01c, possible transposase, partial CDS, len: >  
 150 aa. Similar to many e.g. Methylobacterium sp TR:050409  
 (EMBL: X96995) transposase IS1357 (411 aa), fasta scores  
 opt: 239 z-score: 312.5 E(): 4.7e-10 38.2% identity in  
 131 aa overlap."  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative transposase"  
 /protein\_id="CAB56345.1"  
 /db\_xref="GI:5918467"

Query Match	Best Local Similarity	5.2%; Score 53.6; DB 1; Length 41622;
Matches 185; Conservative	0; Mismatches 219; Indels 0; Gaps 0;	
QY 390	cgcttcacatggtccctctgagatgttgccctcaacctgaggggtcaccggcagacgct 449	/codon_start=1 /transl_table=11 /product="phosphoribosylamine-glycine ligase (EC 6.3.4.13)" /db_xref="GI:5918470" /protein_id="CAB56348.1" /gene="purD" /note="Pfam match to entry PF01071 GARS, phosphoribosylglycinamide synthetase (GARS), score 697.70 E-value 5.5e-206" 5128..5151 /gene="purD" /note="PS00184 Phosphoribosylglycinamide synthetase signature." 5857..7851 /gene="SCD25.05" 5857..7851
Db 39324	CGCCCTCGAGCGGGGCTGAGTGTCTGTGCGCACCCGGCGGCTCGGACATCAT 39383	/note="SCD25.05, possible integral membrane protein, len: 664 aa. Contains a degenerate direct repeat region: residues 469_PYPADDEA P HAPYADDEA L HYPADDEA P HAPYAD 502. Also contains possible membrane spanning hydrophobic domains." /codon_start=1 /transl_table=11 /product="putative integral membrane protein" /protein_id="CAB56349.1" /db_xref="GI:5918471"
QY 450	ccagcctggaagaaggtccttgagaaaaagggcctggtgcccacgcagctccca 509	
Db 39384	CAACCCGCGCGCTGCTCCCTCGAGAACGTGAGATCGCGGTCTTCGACGAGCGCCACA 39443	
QY 510	aaccgtcaacgggagcgccgggccaatcgcaacctttggcgctccggctgaagccag 569	
Db 39444	GATGTCGACGCGGGGCTTCCTCCGCCGAGGTACACCGAGCTCTCGACAGAGTCCCGCCG 39503	
QY 570	gaaagccaggtccaacctgagcagactaaactaacaccttgaggaacctgcccata 629	
Db 39504	CGGCCAGCGGATGCTGTCTTCGCCACCATGGAGAAGCATCAAGACCCCTGTGTGACCG 39563	
QY 630	ggcgaacggcggtctctctcaactgggtcaagccctcaacggagcaacagagatccgc 689	
Db 39564	GTACCTGAAGGACCCGGCCCTCCACAGAGTCAACGCCGCCAGGGTCCGTGACACCAT 39623	
QY 690	caacctgagcgtgtgtctctctgtgctcaaggaggaagagtgatgccaaacccaag 749	
Db 39624	GTTCGACACCATCTGTTGTGAACCCCAAGAGCGCGGTACACCGCGCATCGC 39683	
QY 750	cgtagccgttgacggcggtcattctgtctctcccgaggtg 793	
Db 39684	CTCCCGCAAGGCGCACCATCATCTCTCGCACCCACGACTGG 39727	

LOCUS	AC105260	100772 bp	DNA	linear	HTG 28-DEC-2001
DEFINITION	Oryza sativa chromosome 5 clone OJ115D04, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.				
ACCESSION	AC105260				
VERSION	AC105260.1	GI:17985870			
KEYWORDS	HTG: HTGS-PHASE2.				
SOURCE	Oryza sativa.				
ORGANISM	Oryza sativa.				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.				
AUTHORS	1 (bases 1 to 100772) Chow,T.-Y., Hsing,Y.-I.C., Chen,C.-S., Chen,H.-H., Wu,H.-P., Liu,S.-M., Chao,Y.-F., Chang,S.-J., Chen,T.-R., Chen,Y.-B., Chow,M.-H.,J., Hong,Y.-C., Hsing,J.-N., Hsu,C.-H., Huang,J.-J., Kau,P.-I., Lee,M.-C., Leu,H.-L., Lin,S.-J., Wu,L.-F. and Shaw,J.-F. Oryza sativa BAC OJ115D04 genomic sequence Unpublished 2 (bases 1 to 100772) Chow,T.-Y. and Hsing,Y.-I.C. Direct Submission Submitted (28-DEC-2001) Institute of Botany, Academia Sinica, 128,				

COMMENT
<p>Section 2. Academia Road, Nankang, Taipei 11525, Taiwan</p> <p>The nucleotide sequence of this BAC clone was generated by combining Monsanto and ASPC-Taiwan sequencing data.</p> <p>* NOTE: This is a 'working draft' sequence. It currently</p> <p>* consists of 2 contigs. Gaps between the contigs</p> <p>* are represented as runs of N. The order of the pieces</p> <p>* is believed to be correct as given, however the sizes</p> <p>* of the gaps between them are based on estimates that have</p> <p>* provided by the submittor.</p> <p>* This sequence will be replaced</p> <p>* by the finished sequence as soon as it is available and</p> <p>* the accession number will be preserved.</p> <p>* 1 39588: contig of 39588 bp in length</p> <p>* 39589 39608: gap of unknown length</p> <p>* 39609 100772: contig of 61164 bp in length.</p> <p>* Location/Qualifiers</p> <p>1..100772</p>
<p>FEATURES</p> <p>source</p>

```

BASE COUNT      28103 a  22456 c  22621 g  27572 t    20 others
ORIGIN
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="5"
/clone="OJ115D04"

```

Query Match	5.2%	Score 53.2	DB 2	Length 100772
Best Local Similarity	50.8%	Pred. No. 2	9	
Matches 127	Conservative	0	Mismatches 123	Indels 0
			Gaps	0

QY 543 cccttggcccgctccggtctgagccagggaaagccaggtcacacctgagcagctacatcta 602

Db 88516 CGTCTTCGATGGCCGGGAGCTCCGCCCGGAACGACACCGCATCATGCACTACGGCGTTC 88515

Db 88576 CCACGGCAACGTCCTCCACCTCGTCATCCGCATCTCCGACCTCCGCTCATACCGTCCA 88635

663 ggacctaccaggaaccaggaatccgccccacacctggaacgtgctgtgctcctctggctcaagg 722

Db 88636 GACCGTCCACGGCAAGAATTACAGTTCCGCGTCAGACCCGGCGCCGACACCGTGGCTACGT 88655

Db 88696 CAAGCAGATGCGCAAGAACTCCACCCAGCAGCAGNCCACCCAGACCTCGTCTCCA 88755

QY 783 ccccgagtg 792  
|||||

Db 88756 GCGCGAGGTG 88765

RESULT	10
AP004746/c	
LOCUS	
DEFINITION	
AP004746	167049 bp DNA
Oryza sativa chromosome 8 clone P0035F08,	linear HTG 14-FEB-2002
PROGRESS ***, in ordered pieces.	** SEQUENCING IN

ACCESSION	AP004746	GI:18656392	
VERSION	AP004746.1		
KEYWORDS	HTG; HTGS; PHASE2.		
SOURCE	Oryza sativa (cultivar: Nipponbare)	DNA, clone: P0035F08.	
ORGANISM	Oryza sativa		

Eukaryota: Viridiplantae: Streptophyta, Embryophyta, Archaeophyta, Spermatophyta: Magnoliophyta; Liliopsida; Poales; Poaceae; Erihartoideae; Oryzaceae; Oryza.

REFERENCE  
1 (bases 1 to 167049)  
AUTHORS  
TITLE  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission

**JOURNAL OF PLANT PHYSIOLOGY**  
Submitted (13-FEB-2002) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice genome Research Program, Kannondai  
1-1, Tsukuba, Ibaraki 305-8602, Japan

2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: [tsasakienias.affrc.go.jp](mailto:tsasakienias.affrc.go.jp), URL: <http://rgp.dna.affrc.go.jp/>,  
Tel: 81-298-38-7441, Fax: 81-298-38-7468)

COMMENT  
NOTE: It currently consists of 1 config's Gaps between the config's are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are incorrect.

to be replaced by the finished sequence as soon as it is sequence will be replaced by the finished sequence as soon as it is

\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
available and the accession number will be preserved.

```

FEATURES
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
location/Qualifiers

```

source  
1. 167049  
/organism="Oryza sativa"

[/cultivar="Nippondare"](#)  
[/db\\_xref="taxon:4530"](#)  
[/chromosome="8"](#)

	clone="P0035F08"	
BASE COUNT	45406 a 36605 c 37338 g 47383 t	317 others
ORIGIN		

ORIGIN

Query Match	5.2%;	Score 53.2;	DB 2;	Length 10/045;
Best Local Similarity	46.3%;	Pred. No. 2.7;		
Matches 212;	Conservative 0;	Mismatches 243;	Indels 3;	Gaps 1;

356 gcaaccggagactctgtaggaagtgtagacgctcgtcttcatagttccccctggaatgt 415

Db 166796 GCGCATGGTCTCACCSCGGATGAAGAGACGGCCGAGGCTACTCTGGCGAGAAGTCA 166737

Dh 166736 CCGGAGCTGTGCATCCGTCGCCGACCTACTTCAAGCAGGCGCAGGGCACCAAG 166677  
Qy 416 tggcctcaactggyggtacaccgcagacgcgttcacagccttgtagaaagttccttgaaga 4/5

476 aaagggcctgtgtggccaacgcacgttccttaaccaaacgctcaaggagagcgccgggcca 535

Db 166676 ACGCGGCGTCAATCGCGGGGCTACCGGTGAGACCGCATCATCAACGAGCCACCGCGCCG 166617

Qy 536 tcgcagcccttctggcgccgtcgcgtgagcagggaaagccagcgtccaccctggagact 595

596 acatctacccctgtaggaacctgcgccttagacatgagcgaagcgctgcctcctccaact 635

Db 16656 GCGGCGGCACGTTGAGCGTCAGCATTCCTCGCCATCGACACAGGGCGTGTTCGAGGTCCCTTG 166497

Dy 656 gggctcaaggtccatcaggaaacgaattcg---cccacacctgtagtgcgttcctct 712  
||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 166406 ccccccccccccccccccttttgacccaacgactcatggaccact 166437  
|||||

[illegible]

Db 166436 TCGTCAAGGTCTATCCGCCCGGAGACGCGGCCGCGACATCACCGCGGAGACCGCGCGCGCTGG 166377

QY 773 tccgtctcccccagagtgagcgcttccaactccg 810  
DB 166376 GCAGGCTCCGCCGCGAGTGCAGCGCCAGCGCGCGC 166339

RESULT 11

SC8A2/c 7459 bp DNA linear BCT 09-OCT-2000

LOCUS Streptomyces coelicolor cosmid 8A2.

DEFINITION AL445327

ACCESSION AL445327.1 GI:10798655

VERSION AL445327.1 GI:10798655

KEYWORDS secreted protein.

SOURCE Streptomyces coelicolor A3(2).

ORGANISM Streptomyces coelicolor A3(2)  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 7459)  
Rendenbach, M., Kieser, H.M., Denapalte, D., Eichner, A., Cullum, J.,  
Kinsash, H. and Hopwood, D.A.  
A set of ordered cosmids and a detailed genetic and physical map  
for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
Mol. Microbiol. 21 (1), 77-96 (1996)  
97000351

REFERENCE 2 (bases 1 to 7459)  
Brown, S.P. and Harris, D.  
Unpublished

REFERENCE 3 (bases 1 to 7459)  
Cerdano, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.  
Direct Submission

AUTHORS Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT

Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded  
by the BBSRC and Beaulif Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are  
available on the World Wide Web.  
(URL: <http://www.sanger.ac.uk/projects/S-coelicolor/>) CDS are  
numbered using the following system eg SC7B7.01c. SC (S.  
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary  
strand).  
The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous. The length  
in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for  
CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given  
where these have been used to deduce the initiation codon. Gene  
prediction is based on positional base preference in codons using a  
specially developed Hidden Markov Model (Krogh et al., Nucleic  
Acids Research, 22(22):4768-4778(1994)) and the Frameplot program  
of Bluh et al., Gene 30:157-66(1984) as implemented at  
<http://www.nih.gov.jp/>  
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the  
correct initiation codon. Where possible we choose an initiation  
codon (atg, gtg, ttg or (atc) which is preceded by an upstream  
ribosome binding site sequence (optimally 5-13bp before the  
initiation codon). If this cannot be identified we choose the most  
upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence  
overlapping sections once, or longer, because we arrange for a  
small overlap between neighbouring submissions. Cosmid 8A2 lies  
between and overlaps cosmids 4A7 and C42 on the AseI-C genomic  
restriction fragment.

FEATURES

Source

1..7459  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"

gene  
misc\_feature  
CDS  
gene  
CDS  
misc\_feature  
gene  
misc\_feature  
BRS  
misc\_feature

/clone="cosmid 8A2"  
complement(1..126)  
/gene="SC8A2.01c"  
1..112  
/note="nominal overlap with Streptomyces coelicolor cosmid  
SC4A7"  
complement(1..126)  
/gene="SC8A2.01c"  
/note="SC8A2.01c, possible secreted protein (fragment),  
len: >42 aa. Contains possible N-terminal region signal  
peptide sequence"  
/codon\_start=1  
/transl\_table=11  
/product="putative secreted protein (fragment)"  
/protein\_id="CAC12792.1"  
/db\_xref="GI:10798656"  
/translation="MKRRRIILAGGLVLAALGVCGTGWNTYTQATDGLIANGREND"  
complement(123..1424)  
/gene="SC8A2.02c"  
complement(123..1424)  
/gene="SC8A2.02c"  
/note="SC8A2.02c, possible secreted protein, len: 433 aa;  
similar to TR:053972 (EMBL:AL022073) Mycobacterium  
tuberculosis hypothetical 46.7 kDa protein MTW051.09, 437  
aa; fasta scores: opt: 529 z-score: 488.9 E(): 1e-19;  
31.3% identity in 438 aa overlap. Contains Pfam match to  
entry PF02470 mce. Also contains possible N-terminal  
region signal peptide sequence"  
/codon\_start=1  
/transl\_table=11  
/product="putative secreted protein"  
/protein\_id="CAC12793.1"  
/db\_xref="GI:10798657"  
/translation="MTTLAVRLKNAFLILIALVLSFGIRADGRVYGVADYITVD  
VHLPRGGLFTSHDVYTRGVSVRGVPIDLTAEGVAAELRIKSAIPADPKAVAVG  
LSAAGEOYIDLRPESGSEYLDGTRIBQADVEVPAVYDVLTSVDLASSVPLEDR  
TVDEEGKAFDGHDDLYLIDSGSDVEAADRALPSTLLINDETVLTQAOARA  
IRDPFVAGKIDLAALAKGSPADRLRLAVTPEATOVSGILRLDLSGLVLANLTT  
EYAVTRGRIEELVLYKPAVASGATANGCTLIDGLAFTFSPSLPCPDGVGSTRYN  
GIDLGTAFLNTDACTAPASGCKNVRGSANPKSGAVDPATPSSLPBGGRTPAG  
GSPARDGSTALPGLALPLPGSGEAPAGLTGLAPAGAGAR"  
complement(963..1397)  
/gene="SC8A2.02c"  
/note="Pfam match to entry PF02470 mce, score 98.10,  
E-value 1.8e-25"  
complement(1421..2662)  
/gene="SC8A2.03c"  
complement(1421..2662)  
/gene="SC8A2.03c"  
/note="SC8A2.03c, possible secreted protein, len: 413 aa;  
similar to TR:053971 (EMBL:AL022073) Mycobacterium  
tuberculosis hypothetical 40.2 kDa protein MTW051.08, 377  
aa; fasta scores: opt: 685 z-score: 681.0 E(): 2.1e-30;  
36.0% identity in 311 aa overlap. Contains Pfam match to  
entry PF02470 mce. Also contains possible N-terminal  
region signal peptide sequence"  
/codon\_start=1  
/transl\_table=11  
/product="putative secreted protein"  
/protein\_id="CAC12794.1"  
/db\_xref="GI:10798658"  
/translation="MKRATLPRGRVAGTLAAGLALATLIGSVVPSGPDIE  
DILPGAGADLHPYRVTAELDVLSPVSAVRVDVAVGRITGLGSDVLPATRSRT  
MEINRGVRLPADRTARLEOSSLGEKYVYLVAPAKETGRLGDSVLPATRSRT  
VEINFGALSLILGGVNDLKTITRLNALAGREVERGSMKRVTVLGDIDDRHG  
ITDADLAVRLSLSTLATRKDDVGTITLNLSPGLTERGRGSLTNLRSLDLSGVAV  
SITNAKSDMIADIKAVAPTLRLADAGDLDPSLOVLLTFPTDVLINGVAGDYANT  
YLSMVAVPGETEVIPLVDGDTPEPTASATADQDPDAAKNPAKNPISRRSSGGRGS  
ASPLPLPVSAGAPASGEGH"  
complement(1431..1438)  
/gene="SC8A2.03c"  
complement(2138..2596)  
/gene="SC8A2.03c"

```

gene . /note="Pfam match to entry PF02470 mce, , score 119.10,
E-value 8.5e-32"
complement(2659, .3672)
/gene="SC8A2.04c"
complement(2659, .3672)
/note="SC8A2.04c"
similar to TR:053970 (EMBL:AL022073) Mycobacterium
tuberculosis hypothetical 44.4 kDa protein MTW051.07, 423
aa; fasta scores: opt: 596 z-score: 625.4 E(): 2.6e-27;
33.1% identity in 320 aa overlap. Contains Pfam match to
entry PF02470 mce. Also contains possible N-terminal
region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAC12795.1"
/db_xref="gi:10798659"
/translation="MTRRKILTVGLLVLAAGLAARALPGGRTVAYPEPRAIG
VYASGLRIVGRVGEVSDEPGRVVGRLDDGIKVPEDAAVAAYVAPSVADRIY
VATPLVRRICPALADAVIPASRNHPVPIIDITYSITELGALPEADANAGASELL
RIGADNLONGCEFAIDGVEQFGRKAKRTIDGSSGDFETLSLQFTFMKKKIDVPT
AQRERLVAFFPADKKDDLTGALDELGALQVTFIEDNNGELKKVNDRIYPTFRL
VQORSLEALVADPLADLVNVAYNVNPTRILDRANLNETSISGPIILPEVAOTEG
EGR"
complement(2667, .2671)
/gene="SC8A2.04c"
complement(3238, .3654)
/gene="SC8A2.04c"
/note="Pfam match to entry PF02470 mce, , score 103.70,
E-value 3.6e-27"
complement(3669, .4724)
/gene="SC8A2.05c"
complement(3669, .4724)
/note="SC8A2.05c"
similar to TR:053969 (EMBL:AL022073) Mycobacterium
tuberculosis hypothetical 43.7 kDa protein MTW051.06, 410
aa; fasta scores: opt: 683 z-score: 769.5 E(): 0; 33.4%
identity in 368 aa overlap. Contains Pfam match to entry
PF02470 mce. Also contains possible N-terminal region
signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAC12796.1"
/db_xref="gi:10798660"
/translation="MKPRGKPLFRPKERNPVAVAGALLVTLVALVYVNDRLPEF
GGGVTVADGFSASGLDEGDGVRVAGVGVTVVALDGAVKVSEVEDAWAGIDRT
AAIAIKTVLGRFVIALDPLDGSGRDPPGARIPIDARTSPYDTQAFQDLSGVDDIGT
RLAEFTLIDSPFKSPPHVAKATKATGLDLSKSIKRDGAKISELKSAGKQPIITLNN
KSPFTLIEDGPIGLGELRRRTAIVNMLDLSGSDILGELGIVYDNRKSCGPIIKALG
RVTSVLEKNNRIGETALTALVPTVRLVGNLTGNRWFDSYLCGVPRDYLPETISPT
GCLPKQRAAAGSGAR"
complement(3682, .3685)
/gene="SC8A2.05c"
complement(4254, .4670)
/note="Pfam match to entry PF02470 mce, , score 112.80,
E-value 6.7e-30"
complement(4721, .5785)
/gene="SC8A2.06c"

```

Query Match	5.1%;	Score 52.6;	DB 1;	Length 7459;
Best Local Similarity	46.7%;	Pred. No. 5.5;		
Matches 237;	Conservative 0;	Mismatches 264;	Indels 6;	Gaps 2;

0Y 286 gccgcgcacgaagcggccctccaaagactcttgaaactctccaggaagatgcccgctcac 345  
 Db 983 GACGACCTCGCGAGACTCGTCCGCTCGGAGACACTGGCAGACGTCGTCGACGAGATTGGC 924  
 0Y 346 atccctcctcggaacccggaagctctggaagaagctggagacgctcgatcattcatgctccc 405

Db 923 AAGGCTTGGAGGAGGAGGCGACGACGCTTCAGGTCCTCTGACAGCGGACGACGACTTC 864

QY 406 ctgagagatgttgcgcctcaaccttgggtgtcaaccggcagaaccgttccacgcttgaagaag 465

Db 863 GTTCGAGGCGCGGACGCGCCCTGCGGCTCCACACGCGCTCATCAACGACGCGGAGACC 804

QY 466 gtcccttgagaaaaagggccttgggttgcaccacggcgttccttaccaaacgltcaaggggag 525

Db 803 GTCTGTCGCAACCGACGACGAGGAGGCGGCGGCGCATTCGCGACTTCGCGCTCGCGCGAAG 744

QY 526 cgcgcggccatctgcgaaccttgggcgcgtccgcgtctgaagccaaaggaaagccagcgtacc 585

Db 743 GACGTGGCGCGCGCCCTCAAGGGGCTCCGACGCGCACTCGCGCGCTGCTCCGGTCAAC 684

QY 586 ctggaagcatcatatacctcccttgaaggaaact---cgccctagacatgtgccaaagcgctgt 642

Db 683 CCCGAGGCGCGCACCGAGGTACGCGGCGCTGTCGAGGACCTCGACCCGCGAGCTCGGGGTC 624

QY 643 ctctccttcaactgtgttcaaggcttaccagaagccaaagaaatccgcggcccaaccttgaagtgt 702

Db 623 GTTCTTCGCCAACCTCTGACCACTGCCGAGGTCGCGCTCACCCGCGACGCGGCAATCGAG 564

QY 703 ctgtgtccttgggtctcagagggaagaaggtgtatgtcccaacaacaaagaacgttgcgtgt-- 760

Db 563 GAACTCTCTGTGAAGTAACCGCGGCGCGCTCTCGCGCGGTGCCACGCGCGTCAACGGGGG 504

QY 761 -acctggagcctacatccctgttctctccc 786

Db 503 AAGCTGGAOCTTCGCGCTGCGCTGAC 477

## RESULT 1.2

LOCUS	SCM11	34983 bp	DNA	linear	BCT 01-DEC-1999
DEFINITION	Streptomyces coelicolor cosmid M11.				
ACCESSION	AL133278				
VERSION	AL133278.1	GI:6522830			
KEYWORDS	acetyltransferase; alpha-mannosidase; dehydratase				

## KEYWORDS

## SOURCE

ORGANI

## REFERENCES

**AUTHOR**

3434

FILE

JOURNA

MEDLIN

## REFERENCE

JOURNAL

## REFERENCES

**AUTHOR**  
**TITLE**

JOURNAL

COMMENT

Cerdano, A.M., Parkhill, J., Barrrell, B.G. and Rajandream, M.A. Direct Submission  
Submitted (30-NOV-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
C810 ISA E-mail: barrrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK  
Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at The Sanger Centre are available on the World Wide Web.  
URL: <http://www.sanger.ac.uk/projects/s-coelicolor/>  
CDS are numbered using the following system eg SC78A.01c. SC (S.

CDS are numbered using the following system eg SC7B7.01c. SC (S.

coellicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.

The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nh.gov.jp/jun/cgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation

codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid M11.  
Location/Qualifiers  
1. 34983  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid M11"  
1. 101

## FEATURES

## source

## misc\_feature

## CDS

## CDS

## RBS

## gene

## CDS

/note="nominal overlap with Streptomyces collicolor cosmid STM10"  
complement(1..1764)  
/partial  
/gene="SCM11.01c", hypothetical protein, len: >587 aa;  
similar to various hypothetical proteins, e.g. TR:092B0 (EMBL:AL035161) Streptomyces coelicolor conserved hypothetical protein SC9C7.20 860 aa; fasta scores: opt: 588 z-score: 654.5 E(): 4.4e-29; 30.4% identity in 542 aa overlap  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein SCM11.01c"  
/protein\_id="CAB61912.1"  
/db\_xref="GI:6522831"  
/translation="MSDRGASAPSLPDDWPAHPDITLALNRMSFMDLADLGHMDA QAHVFDLRPEEYDGRPESLALRVPTAESRMDITIVARAKMGSENGTYPRLRRDG TLRWTHQGYTRDETGRRPRTIGIVRDTQEMADIASRQALDEARRRLTVVGL ATALAHARIVVDVLDVLDHGLTRCATSLVAGLVAGRIIRLVADPENSFPGTR VTRIDEPYPMSEAVRTLSRPFIESPEEFAERYPGIMDITHLDTAAAYPLINQAR IGATGILYSDBRHSFSDRNYLVALGSSIASIDRAMIYEDMDLAEGLDQAMLRPTI PSVPGCVAYVYRAASITGALGRDVGMDLPLPGRGVAVIGDVGHDTAAAVM GQULIVLRAIVAEHSPATVVARASVFLHETDRFATCLAEADLGGVQVYVRAH IDPLLRGDSGTCRLRVVEGGLPLGISAEFGTAYLVATLIDPNTLLCTGDVIEP GADLDEMDVLTALITSGPODVRDLADRLIDVDERDDVALVLRHGLGAPRTE GRVOOHVSPGPECLTEARHMI"  
complement(1..1764)  
/gene="SCM11.01c"  
1874..1879  
1877..1879  
1877..1879  
1887..1889  
/gene="SCM11.02"  
/note="SCM11.02", possible integral membrane protein, len: 470 aa; similar to TR:P96418 (EMBL:Z92669) Mycobacterium tuberculosis hypothetical 54.8 kD protein MTCY08D5.31c, 482 aa; fasta scores: opt: 2112 z-score: 2527.5 E(): 0; 61.9% identity in 472 aa overlap and to TR:069633 (EMBL:AL022121) Mycobacterium tuberculosis putative peptide transport system permease; fasta scores: opt: 122

z-score: 149.8 E(): 0.58; 27.7% identity in 166 aa overlap

/codon\_start=1  
/transl\_table=1

/product="putative integral membrane protein"

/protein\_id="CAB61913.1"

/db\_xref="GI:6522832"

/translation="MEMTPADPYWIALRYFQRALAVYLVAFUPLALOPRALGERGL TPVPEFVERPERRRSLFQWRTSDRLRAGCAVSAALAGLDLPLWGMML WLVPALISTLVNGQVWYFSGWESLLETGCAVAVLENEVAPVYVLPFLRPLFLR VEGAGLTKMRDECEMRKLTICDHHETGMPGLSPFNHLPKLIHVEVAANVTQ LVPELPLAPHPVSPAAALMATQLMYLSGNFSLMNVITLVATSVRRPADSV AAAPLVYVAVVALVALVPLFSRPRVMSIRQVMSFOSIHLVNTYVPGSVSR RVEVIECTADEVAKKDDKREYFKGRCGPRRRPQAPAYHLRLDLMLFALSPS YAGSFRGFVERBLENDATTLTKRSSFPDPAPRFVRAFLRFRTYTWBLRETGA CWERTYREYLPRLTAPR"  
complement(3319..6342)  
/gene="SCM11.03c"  
complement(3319..6342)

/note="SCM11.03c", possible alpha-mannosidase, len: 1007

aa; similar to TR:AC00190 (EMBL:AF04414) Homo sapiens

alpha-mannosidase 6A8B, 1062 aa; fasta scores: opt: 1326

z-score: 1446.5 E(): 0; 37.8% identity in 1013 aa overlap

and to SW:MAN1.RAT (EMBL:M57547) Rattus norvegicus

alpha-mannosidase (EC 3.2.1.24), 1040 aa; fasta scores:

opt: 1285 z-score: 1401.8 E(): 0; 35.9% identity in 1041

aa overlap

/codon\_start=1

/transl\_table=1

/product="putative alpha-mannosidase"

/protein\_id="CAB61914.1"

/db\_xref="GI:6522833"

/translation="MHDRSLVEARLKRVLDERVAPALYPESVPLDVAVMNAPGPVP VEEGLAEPRIEVARMGAPWGTSGVTGVTPEMGKVEAIIIDGPEENMGFO CEGLYVRPDTGPKVGLNPNQWNRIGLAVEEGEQVRLVEASNPVILDPFVPTRL GDKDTAGSEPOYTLRMDLAVLDETVMNLVLDLGLGEMLELPVDSRREITLRVD KALDAIDLDVNGTAEGARSRLTQVAPVAPVSRHISAVGHAHIDSMILPEYIK KAVRTSMKTLIDDEPVPYAKQAQWAPVRRHPRVYKAKVADGTFVAGK WVESDTNPGSEBAMROFVHGKREFLDFEGEENDEAMVLPDVFGRAGLPQITKAGK YLTKQISWQTNKFPNHTFMEIGIDRTIFTHPPVDTVNCSSKMGSIANAANFD KVARHSLAPTMGDDGGGTTREVMYAKARLNLESGTVEMETPHAFEEKAEANEA PPVWVGEIYLEHRAITLTSQAKTKOGNRSRHLREALATKATAVRAVGFYPPDDD RIMKTVLLHOFHDILPGSSIAVREARATDYVAALNGLIIDAORLAGEGPTPIY FNSAPHARAGVPAGAASPAIDGTGSGPRGGHYVDNGLIRTEIDRIGIYATIL AADRETIAPGAGNLIQHPFPMMDAMDVETRTNTVDTLDADEVPRQDDASVR ITRSGSSRTQVLTLLAVERRLVDVDEVDHETKEPLKIAFPDVAERVASTQSG HHRPHTNHSWEAKFEACNHRFVHLDEPQVYALVNDSTYGHDTVTVTLDACT TTIVRLSLRAPRPDETTDGGVHRFRHALVPGAGIGDAVEGRINLPEHLTGAP EVAPLVYDRVAVVTVAKLADDSGDVYVFEHAGGRTATTLAGFAVADVDTLL LERPLADEARPPGDRITAVSLRFOULTLRKA"  
complement(6350..6353)  
complement(6484..7731)  
/gene="SCM11.04c"  
complement(6484..7731)  
/gene="SCM11.04c"  
complement(6484..7731)  
/note="SCM11.04c", hypothetical protein, len: 415 aa;  
similar to TR:CAB5687 (EMBL:AL121596) Streptomyces

coelicolor hypothetical 71.4 kD protein SCF51A.35, 644 aa;

fasta scores: opt: 205 z-score: 234.7 E(): 1.1e-05; 26.0%

identity in 408 aa overlap

/codon\_start=1

/transl\_table=1

/product="hypothetical protein SCM11.04c"

/protein\_id="CAB61915.1"

/db\_xref="GI:6522834"

/translation="MPASVRCGVVYTPVSGFHHMLDFDLDSVRADLSDISAALDVHY RVPLMPYFQPNRLIRERAVEDVAVLVDAGEGDLVNVNDLQGHSSPDYVAPWR TWHRRNLFTPDVYGOAALYRTALAAVLGAPNLTGLTINENVOFSGAPPPDRAT SAQIDAMLERNLAACEGAPGRMLIAEYATVYQODOPPTPQAAQCVATVHSHV FNGRAOCHGRISVSEHHAAYLILSTAAKDDPHRPVWDEVGAPALVFAEHAAAT EATVENLDCPEDLMGVTKWCSHDVSRALAPPELGTGLITNDRRKDTARVLAASA RAARDGARPAAPRTTALVADPADBASACAPGAGVADAEFRILVADGARPTVLTGPA

Query Match	Best Local Similarity	Score	DB 1	Length
Matches 157; Conservative	47.4%	5.1%	Pred. No. 4.4;	34983;
	0;	Mismatches 174;	Indels 0;	Gaps 0;
334	gccgcgtccacatccctctaagcagccggagagctcttgaggaagtggtggagcgtgctc	393		
Db 21757	gtccacgtccaccgcccgcgacccctcctaagctgtgcgacgccccgctgcacccccc	21698		
394	ttcaagtcgcccttgagatgtgttgacccctaaccctgggggtcaaccggcagacgtccac	453		
Db 21697	gccggtcttcacgcagaccactgcgaacgggtgacggcgccgcacacgtgcgtggcag	21638		
454	gcttgagaaagagctccttgagaaaaagggcctgtgtgccaaccgacgtcccttaacaaac	513		
Db 21637	accnagccggaccccgccgaccccgagctccctgctgcacccctgcacacccctgctgc	21578		
514	gtcaacggggggagcgccgggcatgagccaccttgtagccgttcggcttgtagggcaaggaa	573		
Db 21577	ggcaacaaaggagcgacgacggccggccacccggcagctccgtacagacggcgaaggcgcc	21518		
574	gccaggtcaacccttgagacgactaaccttaacccttgaggaacctgcgccttagacatggcc	633		
Db 21517	cttcaccgacacccgaagacactgttcacagccttcttgagacagggggcgacagacgtcc	21458		
634	aacggcgctgctcctctaactgggtcaagg	664		
Db 21457	acctcagagctgcgacacgcagctggcgccgacg	21427		
RESULT 13				
SC4A10/c	SC4A10	43147 bp	DNA	Linear
LOCUS	Streptomyces coelicolor cosmid 4A10.			
DEFINITION	Streptomyces coelicolor cosmid 4A10.			
VERSION	ALI09663			
KEYWORDS	ALI09663.1 GI:5689939 carotenoid dehydrogenase; cell division protein; DNA-binding protein; ftsi; ftsq; ftsw; ftsz; isoleucyl-tRNA synthetase; lipoprotein signal peptidase; metz; methyltransferase; reductase; methyltransferase; murD; murF; murG; murX; Na+/H+ antiporter; peptidoglycan biosynthesis; pseudouridine synthase; sporulation protein; thiamin phosphate pyrophosphorylase; transcriptional regulator.			
SOURCE	Streptomyces coelicolor A3(2).			
ORGANISM	Streptomyces coelicolor A3(2)			
REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.			
AUTHORS	1 (bases 1 to 43147) Redenbach, M., Kleser, H. M., Denapate, D., Eichner, A., Cullum, J., Kinsash, H. and Hopwood, D. A.			
TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome			
JOURNAL	Mol. Microbiol. 21 (1), 77-96 (1996)			
MEDLINE	97000351			
REFERENCE	2 (bases 1 to 43147) Saunders, D. C. and Harris, D.			
AUTHORS	Unpublished			
JOURNAL	3 (bases 1 to 43147)			
REFERENCE	James, K. D., Parkhill, J., Barrell, B. G. and Rajandream, M. A.			
AUTHORS				

TITLE Direct Submission  
 JOURNAL Submitted (02-AUG-1999) Streptomyces coelicolor sequencing project,  
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
 Colney, Norwich, Norfolk NR4 7UH, UK  
 COMMENT Notes:  
 Streptomyces coelicolor sequencing at The Sanger Centre is funded  
 by the BBSRC and Beowulf Genomics  
 Details of S. coelicolor sequencing at the Sanger Centre are  
 available on the World Wide Web.  
 (URL: <http://www.sanger.ac.uk/Projects/S-coelicolor/>)  
 CDS are numbered using the following system eg SC7B7.01c. SC (S.  
 coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary  
 strand).  
 The more significant matches with motifs in the PROSITE database  
 are also included but some of these may be fortuitous.  
 The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for  
 CDS which show significant similarity to other CDS in the database.  
 The position of possible ribosome binding site sequences are given  
 where these have been used to deduce the initiation codon.  
 Gene prediction is based on positional base preference in codons  
 using a specially developed Hidden Markov Model (Krogh et al.,  
 Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot  
 program of Bibb et al., Gene 30:157-66(1984) as implemented at  
<http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the  
 correct initiation codon. Where possible we choose an initiation  
 codon (atg, gtg, ttg or (atc) which is preceded by an upstream  
 ribosome binding site sequence (optimally 5-13bp before the  
 initiation codon). If this cannot be identified we choose the most  
 upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the  
 sequenced clone. It may be shorter because we only sequence  
 overlapping sections once, or longer, because we arrange for a  
 small overlap between neighbouring submissions.  
 Cosmid 4A10 lies between and overlaps with cosmids 4G6 and 2C1 on  
 the AseI-c genomic restriction fragment.  
 Location/Qualifiers  
 1..43147  
 /organism="Streptomyces coelicolor A3(2)"  
 /strain="A3(2)"  
 /db\_xref="taxon:100226"  
 /clone="cosmid 4A10"  
 1..102  
 /gene="SCA10.01"  
 /note="overlap with Streptomyces coelicolor cosmid 4G6"  
 1..205  
 /gene="SCA10.01"  
 <1..205  
 /gene="SCA10.01", hypothetical protein, partial CDS, len:  
 >67 aa; unknown function, probable CDS suggested by GC  
 frameplot and positional base preference. Continues in  
 cosmid 4G6 as SCA46.39"  
 /codon\_start=2  
 /transl\_table=1  
 /label="SCA10.01  
 /product="hypothetical protein"  
 /protein\_id="CAB51977.1"  
 /db\_xref="GI:5689940"  
 /translation="TRANRRHRVWDTRDRCGYGLDITADRAQMDYVVSDRTDPRAAI  
 SEWTRSTKSTGTOFVERVYSPA"  
 complement(190..780)  
 /gene="SCA10.02c"  
 complement(190..780)  
 /gene="SCA10.02c"  
 /note="SCA10.02c, hypothetical protein, len: 196 aa;  
 unknown function, probable CDS suggested by GC frameplot,  
 positional base preference and amino acid composition.  
 annotated in cosmid 4G6 as SCA46.38c"  
 /codon\_start=1





SOURCE Streptomyces nodosus.  
ORGANISM Streptomyces nodosus  
Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:  
Actinomycetales: Streptomyces: Streptomycetaceae: Streptomyces.  
REFERENCE 1 (bases 1 to 113193)  
AUTHORS Caffrey, P., Lynch, S.V., Flood, E.M., Finnan, S.M. and O'Leary, M.  
TITLE The amphotericin biosynthetic gene cluster from Streptomyces nodosus  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 113193)  
AUTHORS Caffrey, P., Lynch, S.V., Flood, E.M., Finnan, S.M. and O'Leary, M.  
TITLE Direct Submission  
JOURNAL Submitted (07-MAR-2001) Industrial Microbiology, University College Dublin, Belfield, Dublin, Ireland  
FEATURES  
source 1. 113193  
/organism="Streptomyces nodosus"  
/db\_xref="taxon:40318"  
/complement(4. .1824)  
/gene="amphC"  
/complement(4. .1824)  
/gene="amphG"  
/note="possible amphotericin export protein"  
/codon\_start=-1  
/transl\_table=-1  
/product="AmphG"  
/protein\_id="AAK73498.1"  
/db\_xref="GI:14794890"  
/translation="MAPDVEHEHEEERESROPYRLAAILRPHRSVGATAGVYGI  
LWAFGRILLGRVTLADGCVLHGGRPAVVDGALGRLLMILLYIVASVMLVON  
WLYASVRLILHDLRRAQOKLRLPLRFDKRPACETLISRGTDVDNLOQTLLT  
DLISVSLVIMLSLMLIISPLAGVRLSVPSGLLAMIISRAQOPYAAMASAGK  
LTAHEEMCGHALVKAIFRRAEAEQFDENEAVYRAGSAQFASGALPEVMMFAN  
LGVAAYVAGAMKVNKSLIGDQAFIYARQSPQIYVIAVAGRLSGVSAQV  
FTLLDPEOEPEPRDLAVERGEVPEODVSRFRRYRDPDLRSGVLDPTGKTI  
GPGACKTAYANLMREYETDSGRILLDGPDTAMNRDLSRSGVLDPTGKTI  
AENIATGSRATRAQVYEAARATYADRFTITLSQIDTYLDDESGVSGEKOITVA  
RAPLARPAVLDDEATSVDRTELILQRMNLTGRGRTSFVLAHRLSTIRDAIVV  
MKSGRIVEGTHOQLIDAGCAVYRLHAARADAPADVTG"  
gene  
complement(1805. .3628)  
/gene="amphH"  
/complement(1805. .3628)  
/gene="amphH"  
/note="possible amphotericin export protein"  
/codon\_start=-1  
/transl\_table=-1  
/product="AmphH"  
/protein\_id="AAK73499.1"  
/db\_xref="GI:14794891"  
/translation="MAPSVRWPTTFFPORAAGRQPCGSVLRIMTQLRPVAGSYA  
LVVLAHVLIIGLILPLTGLAALIDEGVVRHDSRIGTGTAKVAVVALVOIAALGAA  
LGAARTSTALGRDLRSVFRVLDPSAREGTPSLITRSVNDVOQONLAQSGGI  
FVAAPLMCGSVLLARQDVTALILIPVLYVAVCGILLSRMALVYRLQOTLRI  
GALLRRTGVRVRSFADHGERFTTNELGLSGVGRILVAMPRLVLLMNL  
FTLGLVWGARRIDSGMOIGALSAFLSYLSILMSVMAPEVLVAMPARVAREIT  
EVQADTDVPPASPRPMAGPGVGLGAEFRYPAENAVLRLDLSILTRPGRVAVL  
GSGSGKTTLLILRLVDVTAGVATIGGDVRELDVSWAAVAVGVPPVYEFATV  
ASNMFRGRPDATDEELWEVLRLAQHGEFVTRGLGDTLTAQGGTIVSGQORRLATAR  
ALRLPAIYLFDDSFALDOSTEAALRKALVPTGEGATVTVQAQVASYRADRIYLL  
DQGGIATGTHDLRDSPTYREIALSQRTRETAAGAGRS"  
gene  
3840. .4874  
/gene="amphDIII"  
3840. .4874  
/gene="amphDIII"  
/note="mycosamine biosynthesis protein"  
/codon\_start=-1  
/transl\_table=-1  
/product="AmphDIII"  
/protein\_id="AAK73500.1"  
/db\_xref="GI:14794892"  
/translation="MPKRALITGITGQDGSYLAENHLSQGYQVWGLIRGOANPRKFRV  
SRLASLSEVDGLDQGLSVSAVDKPEVNLGALISFVPMKSMQGAELVTEVNGM  
YLAVLEAIRWVSGLSMSTRAGTEGQIRFYQASSSEMGVAAETPPQRETTLLPHRPSYG

AAKAYGHFTRNRESFGMYAVSGMLENNHSPRQGEPTYRKLISLAVARKLQDLKL  
ALGNMDVPRMGVAGYVYAMHMLQDDAPDVVITGEMHYRDAVRAPAEVYGDW  
KQYVVVDPLVRAEVEVLCADSKKAQOLGKRPVDFQELNRMNMDADLASVRQNE  
LDDLALHSW"  
5042. .33574  
/gene="amphI"  
5042. .33574  
/gene="amphI"  
/note="polyketide synthase multienzyme polypeptide housing  
extension modules 9, 10, 11, 12, 13, and 14"  
/codon\_start=-1  
/transl\_table=-1  
/product="AmphI"  
/protein\_id="AAK73501.1"  
/db\_xref="GI:14794893"  
/translation="MDNEOKLRDYIKLATRADLRARRRVELESASQEPATVGMTCR  
YFGVSSPEDLRMVEAENGVTPEFTDGMDEALASSPTSGGTLADPPEADAF  
GISPREAVAMDPOQRYVLESAMEAFERAGINPTSVKSGRTGVEIGMADORYPADG  
AEGFQULTGNTGVSLSGRISYTFGTGPAVTVDPACSSSLVAHALQALRAGCTIAL  
AGGVTVMSGPTFIEMRGOGSLVDRCSRSGFTADGTGMAEGVGLIVERLSDAIRN  
GREILAVRGTAVNODGASNGLTPANPQSOQVIEQALVNARLSAGDIDVEAHGCT  
TIGDPEAQAALLATVGOORDEKPIILGYSKSNISHTQAAACVAGYIKKMMRHCVL  
PRLTLDDEPTRHNDMSQAVRVLTEMTTEMPATGAPRRAVSSFGISGTAAHTIYEDAP  
EPEPADPEDDAPSTPAVTVGLPVLISGSPSEYLRQAQALITTLGTGTPPADLAY  
SLATRTAEHRRAVLASDLPELTGRITAIABETDPVALADVTGTARTETRLAVLT  
GGOAORLTGARELTAAREPVFAALDALDALDPFLHDVPLRKYLMGSDADRLRTEYAO  
PALFAVEVALYRLSEFEVKPDHLASGELIAAAHVAVFSISDADATVLAARGRLMO  
ALPEGGAAMAVOASDEEVAPLAGHEDVLSIAAVNPSVLYSGDPTTETLAARLAA  
DGRKTSRLVSAFHSPPLAAPHLDERRNYEGITLSPILPVVSDTGTGPAVQOITS  
PDWVHVQAVNAFADGIDMLAHADVTALREDPVLSAMQONCLDAGSDALTPA  
LREGREDHTFALALAHPLHOGTALHMDACFTGTGRARDLPYTFQRRYVPRAYOG  
GADLDSVGLIGAHNPLLSAAYSIADESEGLITGRISLSHPMLADHYRTAVLPGT  
AFLEILVRRAGEVGCVRDELTLAAPLVPEOGGVOLVIMGNPAGSGRSTVGGYLF  
DDEDEPMTSHATGVSASRTSPDATVMPADARTLPEVDGLYERLAGEGCGYLF  
OCGRAMWRGDEVFAEVPLPESGHTDAESGHLRALDSALHAASVVDDEERAAGLP  
FSWEGVSLHAGSTTLTRKLRAPAGATAVIAAADDGQVLSDDLTAHVRAREIDA  
AALVVDALFRLDWEPVTVVAAAGTAAVAGDEPFLRALPOGDLAVPRDLALAA  
DGAVPDTVLLPLGTGPDADPVTAHRAATEDEALAAVTELEDDEERASRLALVIRGA  
TGHDDPAAAVMGVIVSAOSENGRELVLDDADODTPALPAAALTESEPQALVAGEE  
LRAARLVRRPASTAEAVPAFEGGAVLVGTGGDGLAVALRHVAHGVREILVSR  
GGAAAGAAELVAELAESGARATVYACDVMDRAAVAEVLAHVAHVSAGVLDGMAV  
GTTTPERLTIVLRPKYDAMNMLHEARTDILKAFVLFSSVAVTSGPCQANVAGNAF  
LDLALAHRRRAAGLPGLSLMGPEBQGTGGTGLSEDDLRKMAAGTLPALTVDGALL  
DALDLDDDAALPVRDLISVLRAGQEVPEPLRLRGRSRAAVASAGAGIAGRLA  
RLDAESRDELVIDLVRGQVALVIGHATGEVIAAGRAFRLEGDSITLAVELRNLNVT  
GLRLPATVLEVDYPTVSHLASVYLDDELLEGEVAAEVYQRTAAVADDPYIYGMACQPV  
GGVTFSEDMRLVTTBGTDAVSGFPVNRGDVENVLPDDDHGTATYTRSGGFLHEGP  
PDGPFSGMSPREALATDSOORILLIAPASMAIRAGDPVGLGSAAGACVAGYMSGS  
AMGSPPEFGPGSGSSPSLASGRSYTTLGEPATVUTACSSSLVAMRMKQALRS  
QSDATIRNGHQLIAVVGSAVNODGASNGLTAPRGPOQVIRQALASGSLASGDVNV  
EAGHGTCTTGQIDLEAQLLATYGRDPEQPLILGSKVKNIGHTQAAAGVAVIKVM  
SMHGVLPPTLHNDPASHVMDTEGAEVELTQTAPEVGRRRRAVSSFGLSGTVH  
TVLEQAPGTTVAAPERTAGVAPLLISGSRIDALRAQAAALTLHONHPPSLADL  
GHSLATRTREPRERAAVITQODREGILASGLSAGSAPPAVVEGEAAGARAVAFMS  
QOSQRAKMERRELETPRPAAPDEVCALDLDLDPDELEVFAEGSEBALDLBGTG  
WQOPALEAVENALRYLVESWGRADVLTGSGEISGAIAAHAVFTLQDARVAAART  
LMEALPESGAMAVAOATEEBEVAVLLEGEISGAIAAVNPVSVVSGEDPVEIETAESG  
RORRTKRLRVSAHSEPHNDAMIDARTYAEITLSVAAPIPIVYSIDTGRADAEVAT  
ADYVWRHVEAEVAFACDVCRTLDAQATLTLETLEGPGLITLAMEADLTGDRVHNHALY  
PLLRADPREESAATAARLQIHGVDIMTYATLACGARGVDLPYTAPOHANYWQPL  
SAAPSAPGADQKLMAAVERGDPAELAVNIDDESILPDLSDLALSAHSMRGNOK  
ALLDITLRYREVENTRISKPAAPVLDGEMVLSSDANADTELDDLEALSAMRGNOK  
RLVLADACADRAVLGARLADTENADATAVLVLPDEPRTGPGAGTGLATLITALY  
QALADTGAAGRILMTATRGAVSTGPADPVPYPOATVAMGRCVATLEHPRILMGGLDLP  
ADPDRGQORLAELVALVADKAPGDEOVVALRAGVARGRLVRIHIVDELPSADQFTVSGS  
VLTGTGGLGAETARMLARSGAHLVLSRSGCAVARGAARLAELEFQSGASVSTVAC  
DYADRDLAAVLDGSLASADPLTGCVHTTAGVGRGPDALITPREAGLTAKLAGAHL  
DNLGDRELDFFTLFGSLAGVSGQOQSVYGAANVLDLALVARRAGDVAISIAMGP  
WGTGMAADAVASGTLRQGLGLDPAALPDLTMRKRAVYQDVTYVAVADVDMVAPLF  
TSARPSALISDLEPVVALAENMTPADVTGASITVYRVSLSPEQDLRLTLTDVITREA  
TVLGHSSAACVADLEPRAFEITGDSITLVAELRKLHAGATLSLSPVMPVFPYPLPLEAO  
YLRAENVGSVLEAVAGVATAGTDDDEIATIGMSCTRPVGSBPEQMLDVLVSGTAIT

[illegible]

```
CDS      73..909
          /gene="RSC2802"
          /function="miscellaneous; hypothetical/global homology"
```

Gene name confidence : hypothetical  
predicted by Codon-usage  
predicted by Homology

```
/codon_start=1
/transl_table=11
/evidence-not_experimental
```

```

/protein_id="CAD16509.1"
/db_xref="GI:17429825"
/translation="MTALLGSAWVCVGNSEAKTLPALGAAGTVYRITTCAILLA

```

SRRAIDPEVMJGAVAGCILLITFVGGRAQOIDEPLGAAYALGAGVCMAJYIVYKRCVGL  
PAGQATSIGMAVGAALPAIFGVQAGPALLAPILYAGLIGVGLSSAVPISLEMAVLR  
HLSGRTFESVLLSEPAIGALGAVLHEHLSRQWAVAIATIAASHCAATPARSRRA

gene complement(936:2024)  
/gene="RSC2803"  
/note="RS00302"

```

/gene="Rsc2803"
/function="miscellaneous: hypothetical/partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CADI6510.1"
/db_xref="GI:17429826"
/translation="MAIEISFEAQYDQARFLDAAARGLAIEPAVPHALGPAG
EPLSDIACFLPDRGALVLTGSHGEGCGSCGVGLRDALFARLAAAGVALL
LHAANVYGFARLRYVEDNVDLNRNSADFAVAANPAIYEDVLLPTWPPDAN
QALLOYLAMHGEGALRDAMTICQYAIIPDMFGGCGTCTNOACATILSRHADAR
LAWIDLTGLGAHGEKITSADPFLERATIGWADVRLIAAGSVSSVEALDF
RACALPFAVKYVITTEFTLEPMAVQALRADHWLHRHPTSPDQAAAIKGLRDAF
YCDTPAWKGVYQARVAVLQAAAFESG"
/complement(2124..3326)
/gene="nrdb"
/notes="Rsc2804: RS00301"
/complement(2124..3326)
/gene="nrdb"
/EC_number="1.17.4.1"
/function="small molecule metabolism: central intermediary
metabolism: 2-prime-deoxyribonucleotide metabolism"
/notes="Product confidence : putative
Gene name confidence : putative
Predicted by Codon_usage
Predicted by Homology"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE TRANSMEMBRANE RIBONUCLEOSIDE REDUCTASE
(SMALL CHAIN) OXIDOREDUCTASE PROTEIN"
/protein_id="CADI6511.1"
/db_xref="GI:17429827"
/translation="MLSMDDVKAPOAPOPAYOPOPILOSPFPPDGGALPSAT
QAGGILGNPNNAASSRNVNADKRVNGATDVNOIYPFKRYKMAEKLACANIN
MPOIINSRDIALMKDPNGUTEDERRITIKRNLGFTYADSLAANVILGYRQTABE
CROLYLQAFEEAIIHTHAQYIVESLGLNAEITFNAYHEVQSIROKDEFLIPIDYLT
DPSEKGTGPENDKLLKSLIVFACIMGELFFVFTOILAMGRQNKMTGAEOYQYIL
RDESLHCNFGIDILNQIKLENPMLWTAEKAEITELFKKADLEYRAEDTMRGVYG
LNAPEKGYLRFICNRCCQIGDALFPEDNPFPMSEMIIDLKKEKNFETRYIEVO
TGGLSMD"
/complement(3477..6440)
/gene="nrda"
/notes="Rsc2805: RS00300"
/complement(3477..6440)
/gene="nrda"
/EC_number="1.17.4.1"
/function="small molecule metabolism: central intermediary
metabolism: 2-prime-deoxyribonucleotide metabolism"
/notes="Product confidence : putative
Gene name confidence : putative
Predicted by Codon_usage
Predicted by Homology"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE RIBONUCLEOSIDE REDUCTASE 1 (LARGE
CHAIN) OXIDOREDUCTASE PROTEIN"
/protein_id="CADI6512.1"
/db_xref="GI:17429828"
/translation="MQTAQTEIHSSTANPTAFSQPAEGTAVASAGYADYKILRRNCA
VVGFEPSKIVAMTKAFILAVNGGASASAVRELEGTQNVNVALILSRSPNGSTFII
EVODQVEILMRSGEHDAVAVYLYRERSORAAOGETQOOPAFIGLNVGGTIR
PLMALRLNIVASACAGISVDPPEPLKTKVTKIKYEGVMSQVYSAIIAATLIEK
DPYQSQTARILHTITREITLGEVTOAEAGNADYFPPIFRGINAGLDDTLAOF
DLAKLGAALDASRDLPFNVLGTLTYDRFLHISDKRIEMPQAFPMRVAMGLSLEID

```

```

REARATFYLISFDEMSSTPTLFNSGTCRSOLSSCYLTTVSDDLDGIYEALKENAL
LSKPFAGIGNDMTNVRALSGHRRGTNRKSGGVYPRFKVNDTVAAVNOGGRGAVAY
YLETHWHDIEEFLERKNTGDDRRITDANTANIPILPFKRVHEGENTLFSPTCP
DLHKVKAFAEATLGEDEKVARGEILFRKMPALQJLPMKMLGLFETGHPWTFKDP
CNIRSPQAHGVHSSNLCETITLNTNDSITAVCNLSVLAHVAVQAQGSYALHD
KLKRTVTRAMRLDNVINDINYAVKKARBDLHRPVGLGIMGFODALHVLRIYAD
AAVOADTSMCAVCYAYMASTELAEGRYSYKSLMORGVLPODSLKLQERGG
YLEADLSSTMDMGLRERIOHGRNRCVVAIPATISNIGVSACIEPTYONLYVK
SNLSGEPTVNDYLRDLKARGLMDPVMADLYFPQSLARIRIPDLIDYATAFE
VEROMLYEASRRKQWIDQOSLNTIYAGSGKMLDDTYRLAMIRGLKITYYLRITGA
THVEKSTVSGTILNAVSSGSDASVAAAGAPVAPASALDTVAATAPAMPDAGAVCT
MRPDGPFCECCQ"
/complement(6877..7479)
/gene="ampd"
/notes="Rsc2806: RS00299"
/complement(6877..7479)
/gene="ampd"
/function="cell processes: protection responses;
detoxification"
/notes="Product confidence : putative
Predicted by Codon_usage
Predicted by Homology"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE AMIDASE PROTEIN"
/protein_id="CADI6513.1"
/db_xref="GI:17429829"
/translation="MEPVAGPVRIGADGMDGVCRRHSPVNDAREGLPIDLIVLNI
SLPPARGTDVLDLDEFANPLDCGAPYEQLRGVSAHFRIIRTGACVGFACDAA
MHAGVDFEGRTRGNDSTGIEGTDOFTPAOVAATAVRAICAAVPIAIGH
SDIAPGRKTPDGCPEFWAHLRLAKLEALFYPQAGAAPA"
/complement(7472..9184)
/gene="pehr"
/notes="Rsc2807: RS00298"
/complement(7472..9184)
/gene="pehr"
/function="small molecule metabolism: global functions;
global regulatory functions"
/notes="Product confidence : probable
Predicted by Codon_usage
Predicted by Homology"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE RESPONSE REGULATOR TRANSCRIPTION
REGULATOR PROTEIN"
/protein_id="CADI6514.1"
/db_xref="GI:17429830"
/translation="MSRQPFPRPMSKAAYRREPLVYDDEADLRELLTSLRMGHD
VYIAGLAEAREALTRORFALVYTDKMLGGLGCLIDYRQLSATRADRPVAVITAYGA
ENAVEALKAGAFDITAKPLSLDQIRSLVYLAALGQODPDGSSADLAERTSALUPGS
AAMQEVRSLLRLARSMAPVYIGSESGSSGARRAARIALUSSARFPAVNCGLAIE
NLMEAEFFGVKAGFTGADSDRGFSPAAHGGLTMDVLDLITMOKVILKRLAQER
VRKIGSEREDPVVRVYVYCAHONLARVLAAGREDIFLYLANVLEMLPRLAEAY
PVLAGVILBOLARVYGLRPRKRLTROLAOLGAPFPGVNRLEIDLELRYAAREGS
IDVDHICALSTDIERSPLFRHAREAHGHPVRAHLRPPVAPRPHAPRPGHVAHPL
GVPOVGGMPDAITIVPVPGLMPHPVAVYFVERPAVTEPAPARPLDVAADGPA
YLSEVERGVIILAAIQTGFNRTASKLILGSLFRQDLRMRQOILIDRPRDILEAAVNGG

```

Query Match 5.1k; Score 52.2; DB 1; Length 188050;

Best Local Similarity 47.8k; Pred. No. 4;

Matches 187; Conservative 0; Mismatches 198; Indels 6; Gaps 1;

```

QY 387 gtcgtcttcattggtcccccgtgagatgttgcctcaacctgggggtcaccgcgcagac 446
      || || | || | || | || | || | || | || | || | || | || | || | || |
Db 140037 GCGCGGCGATGCCG67CAACATGCGCACGACGCGCAGGCGCTACACCGATCTCAA 140096
      || || | || | || | || | || | || | || | || | || | || | || | || |
QY 447 cgtccacgcctggagaagaagctcttgagaaaagcctgtgtgcaccgcagcgtcctca 506
      || || | || | || | || | || | || | || | || | || | || | || | || |
Db 140097 CTTCTCATCCCGGAACCTGTCACGCGGCGATGCGCTACAAAGAAAGCACCTTCTCGAGA 140156

```

OY 507 ccaaacgltcaacgagcgccgggcatcggaaccctltggccgtccggtgaagcc 566  
DB 140157 GGAAGCGACTTCTCGGCGCGCGCGCTCCGATCGACTACGGCGAGACGCTGCCGA 140216  
OY 567 agggaaagccaagctcaacctgagcaactaactaacccttgaggaacctgcctaga 626  
DB 140217 CGGCTTCGCCCGCTCGAGCTGGGGAATACGGCTACCGCGCGCTGCTCGCGACTC 140276  
OY 627 catggccaacgagcgtgctctccttcaactggltcaagcgctacaggaaccaagaaatccg 686  
DB 140277 GCGCGCGTCCGGCGCGGACGCT-----GCTGTACGGCTTCGAGTGGCTGGCGAGNA 140330  
OY 687 ccccaacctgagcgtgctctctggtcctcaaggggaaaggtgatgccacaaccaa 746  
DB 140331 CGGCCCGTGGGTCTGCGCGGAGGCGGTGCACAAGCTCAACGGCGTGTGCTACACCT 140390  
OY 747 gacggtggcgtltgaactgggacctatcctg 777  
DB 140391 GCCGCTGGGCGGCGGAGCGGCTCACCGTG 140421

Search completed: May 29, 2002, 10:21:27  
Job time: 7259 sec

**THIS PAGE BLANK (USPTO)**



PI Wayne J, Xu S;  
XX WPI: 2001-298939/31.  
DR P-PSDB: AAE01000.  
XX

PT Cloning *Thermus* species (Ts) plasmid genes comprises transforming  
PT *Escherichia coli* with cloned recombinant plasmid containing Ts and  
PT *E. coli* origins of replication, isolating cloned recombinant plasmid  
PT from *E. coli* and transforming Ts cell -

XX  
PS Example 1; Fig 1; 32pp; English.

CC The present DNA sequence encodes *Thermus* sp. replication protein, Rept  
CC which is obtained from pTPSP45S plasmid. The replication protein, Rept is  
CC needed for thermophilic plasmid replication.

CC The invention relates to *Thermus* sp. replication protein Rept, partition  
CC protein Para and their corresponding DNA molecules which relates to  
CC recombinant DNA molecules encoding plasmid DNA replication origins in  
CC *Thermus*, as well as to shuttle vectors which contain the same. The  
CC invention also relates to method useful for cloning *Thermus* sp. plasmid  
CC genes which comprises inserting plasmid DNA comprising a *Thermus* sp.  
CC origin of replication (Ori) into a recombinant plasmid comprising a  
CC thermostable kanamycin-resistance gene and an *Escherichia coli* Ori, to  
CC produce a cloned recombinant plasmid. This cloned recombinant plasmid is  
CC transformed with an *E. coli* host cell, and *E. coli* host cell cultured  
CC for the expression of cloned recombinant plasmid. The cloned recombinant  
CC plasmid isolated from *E. coli* host cell is then transformed with  
CC *Thermus* sp. host cell and *Thermus* sp. host cell is cultured. Thus  
CC *Thermus* sp. plasmid genes are cloned. These plasmid DNAs are used for  
CC thermophilic transformation.

CC  
XX Sequence 1026 BP: 221 A; 323 C; 285 G; 197 T; 0 other;

Query Match Best Local Similarity 100.0%; Score 1026; DB 22; Length 1026;  
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaagaacgaacaaacctcttctgaagagctttacgaagctttagagaacccacgac 60  
DB 1 gtaagaacgaacaaacctcttctgaagagctttacgaagctttagagaacccacgac 60  
QY 61 aacacccgtgcacataggggtcagataggggtcagaagaccttcttggccacgac 120  
DB 61 aacacccgtgcacataggggtcagataggggtcagaagaccttcttggccacgac 120  
QY 121 ccccttcagataggggtgcggaacatgccttcggaagagctttacatcaacaaagag 180  
DB 121 ccccttcagataggggtgcggaacatgccttcggaagagctttacatcaacaaagag 180  
QY 181 gcaacttagatgtcttaccggaagaacatagagcttctctctctgttggggcc 240  
DB 181 gcaacttagatgtcttaccggaagaacatagagcttctctctctgttggggcc 240  
QY 241 cccctataccacagcgtgaaccccccggttggaatgatatctcaagccgtccagagcg 300  
DB 241 cccctataccacagcgtgaaccccccggttggaatgatatctcaagccgtccagagcg 300  
QY 301 cctcagaagctcttgaaatcctcgaagagatgcccgcctcacaacatccctacggcaac 360  
DB 301 cctcagaagctcttgaaatcctcgaagagatgcccgcctcacaacatccctacggcaac 360  
QY 361 cggagagctctggaagagtgaggagcgctcctcattcattggtcccttgagagtgtggcc 420  
DB 361 cggagagctctggaagagtgaggagcgctcctcattcattggtcccttgagagtgtggcc 420  
QY 421 ctcaacccgtgggtgcacccgacagacgctcaccgcttgaaagaggtctcttgagaaaaag 480  
DB 421 ctcaacccgtgggtgcacccgacagacgctcaccgcttgaaagaggtctcttgagaaaaag 480  
QY 481 ggcctggtgagccacgacgtctcctcacaacacgctcaacggggggcgccgggcatcgcc 540  
DB 481 ggcctggtgagccacgacgtctcctcacaacacgctcaacggggggcgccgggcatcgcc 540

QY 541 aaccttggccgctccgctgagggccaggaagccaggtctacaccttgagacatac 600  
DB 541 aaccttggccgctccgctgagggccaggaagccaggtctacaccttgagacatac 600  
QY 601 tacacctgaggaacctgcgccttagacatggaacagcgctgctctcttaactgggc 660  
DB 601 tacacctgaggaacctgcgccttagacatggaacagcgctgctctcttaactgggc 660  
QY 661 aagcctacacgaagacacggaatccggccacaccttgagcgtgtgtctcttggtcag 720  
DB 661 aagcctacacgaagacacggaatccggccacaccttgagcgtgtgtctcttggtcag 720  
QY 721 gggaaagggatggtcccaacacgaacggtgaccttgaccttgagcctatcctgtc 780  
DB 721 gggaaagggatggtcccaacacgaacggtgaccttgaccttgagcctatcctgtc 780  
QY 781 ctcccgaggtgagagcgttccaaactccggcccttacacacctatgtctagacatt 840  
DB 781 ctcccgaggtgagagcgttccaaactccggcccttacacacctatgtctagacatt 840  
QY 841 gccgatcctcctagatgacgctgcttcaagaagcttctatgaagcttctgtggctgtg 900  
DB 841 gccgatcctcctagatgacgctgcttcaagaagcttctatgaagcttctgtggctgtg 900  
QY 901 gccaggggtgaactcccgccgaatatctattgacctcccaatggcggtatccgaagat 960  
DB 901 gccaggggtgaactcccgccgaatatctattgacctcccaatggcggtatccgaagat 960  
QY 961 tacacgagtgccatctgacacgacgggagcgtaacctagtgaagacccctcaagagcc 1020  
DB 961 tacacgagtgccatctgacacgacgggagcgtaacctagtgaagacccctcaagagcc 1020

RESULT 2  
AAD04668/c  
ID AAD04668 standard; DNA: 5849 BP.  
XX  
XX AAD04668;  
AC  
XX  
XX 04-JUL-2001 (first entry)  
XX  
XX *Thermus* plasmid pTPSP45S DNA sequence.  
DE  
XX  
XX Replication protein; Rept; partition protein; Para; pTPSP45S plasmid;  
KW kanamycin-resistance gene; thermophilic transformation; Ori;  
KW replication origin; ds.  
XX  
XX *Thermus* sp.  
XX  
XX US6207377-B1.  
PN  
XX  
XX 27-MAR-2001.  
PD  
XX  
XX 14-AUG-1998; 98US-0134246.  
PF  
XX  
XX 14-AUG-1998; 98US-0134246.  
PR  
XX  
XX (NEMO) NEW ENGLAND BIOLABS, INC.  
PA  
XX  
XX Wayne J, Xu S;  
PI  
XX  
XX WPI: 2001-298939/31.  
DR  
XX  
XX Cloning *Thermus* species (Ts) plasmid genes comprises transforming  
PT *Escherichia coli* with cloned recombinant plasmid containing Ts and  
PT *E. coli* origins of replication, isolating cloned recombinant plasmid  
PT from *E. coli* and transforming Ts cell -



PS Example 1; Fig 3; 32pp; English.

CC The present sequence is *Thermus* plasmid pTsp45S DNA. The open reading  
 CC frame of pTsp45S plasmid is the replication protein Rept, partition  
 CC for thermophilic plasmid replication.  
 CC The invention relates to *Thermus* sp. replication protein Rept, partition  
 CC protein Para and their corresponding DNA molecules which relates to  
 CC recombinant DNA molecules encoding plasmid DNA replication origins in  
 CC *Thermus*, as well as to shuttle vectors which contain the same. The  
 CC invention also relates to method useful for cloning *Thermus* sp. plasmid  
 CC genes which comprises inserting plasmid DNA comprising a *Thermus* sp.  
 CC origin of replication (Ori) into a recombinant plasmid comprising a  
 CC thermostable kanamycin-resistance gene and an *Escherichia coli* Ori, to  
 CC produce a cloned recombinant plasmid. This cloned recombinant plasmid is  
 CC transformed with an *E. coli* host cell, and *E. coli* host cell cultured  
 CC for the expression of cloned recombinant plasmid. The cloned recombinant  
 CC plasmid isolated from *E. coli* host cell is then transformed with  
 CC *Thermus* sp. host cell and *Thermus* sp. host cell is cultured. Thus  
 CC *Thermus* sp. plasmid genes are cloned. These plasmid DNAs are used for  
 CC thermophilic transformation.

Sequence 5849 BP; 1377 A; 1501 C; 1670 G; 1301 T; 0 other;

Query Match 100.0%; Score 1026; DB 22; Length 5849;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-230;  
 Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtagaagaacgaaacacccctcttgaagaagctttagaagctttagaagaaacccacgac 60  
 DB 4157 GTGAAGAAAGAAAAAATCTTTGAAGAGCTTTAGAGAGCTTTAGAGAAACCCACGAC 4098  
 QY 61 aacacccgatgacacatgaggggttgaagtagaggggtcagaagactcttcttgccaccgac 120  
 DB 4097 AACACCGATGCCATGAGGGGATGATGAGGGGCTCAGAGACTCTTCTTGCCACCGAC 4038  
 QY 121 ccccccacagatgaggtgcccgaacatcgctgcgaagggcttaccataacccaagaag 180  
 DB 4037 CCCCCCAGATGAGGTGCGCAAAATCGCTGCGAAGGCGCTTACATACCAAAAGAG 3978  
 QY 181 gcaactaagattgcttaccacgaagaagacatgaggtcttctctctctgttgggccc 240  
 DB 3977 GCACCTAGATTGCTTTACCCGAGAAAGACATGAGGCTTCTTCCCTGTTGGGGCC 3918  
 QY 241 ccccccataacacacgacgtacacccccctgttggaaatgataatgtaacgctccagagag 300  
 DB 3917 CCCCCATACACACACCTGAACCCCGTGGGATGATGTAAGCCGTCAGAGAGGG 3858  
 QY 301 cctcagaagcttctggaactcctcagaagattgcccgcgtccacatccctcagacgaac 360  
 DB 3857 CCTCAAGAGCTTGTGAACCTCTCCAGAGATTGCCCGCTCCACCATCCCTACGGCAAC 3798  
 QY 361 cggagagctctggaagaagtgaggacgtctgtctcatagttcccccctggaagtgtggcc 420  
 DB 3797 CGGAGACTCTGAGAAAGGGGAGGAGGCTCTCTCATGCTGCCCTGAGAGATTGGCC 3738  
 QY 421 cttaaactgagggttccacccggcagacggttccacgctctggaagaaggtccttggaagaag 480  
 DB 3737 CTCAACCTGGGGGTCAACCCGGCAGACCGCTCAAGCTGGAAGAGGCTCTTGAGAAAAAG 3678  
 QY 481 ggcctgtgacacgacgactcttcaacaaacgctcaacggagagcgccggccatcgac 540  
 DB 3677 GGCTGTGTGGCCACGACGCTCTTCAACAAACGCTCAACGGGAGGCGCCGGCCATCGGC 3618  
 QY 541 acccttggcgctccggtctgagcagggaaagcagagctcacccctggaagctacatc 600  
 DB 3617 ACCCTTGGGCGCTCCGGCTGAGGCGCAGGAAAGCCAGGCTCACCCCTGAGCAGCTACATC 3558  
 QY 601 taaccctggaggaacgtcgccctagaacatgccaagcggtgtcttctcctcaatggtgc 660  
 DB 3557 TACCCCTGGAGGAACCTCGCCCTAGACATGCCAACGGCGTGTCTCTTCAACTGGGTC 3498  
 QY 661 aagcgctacaggaacacgaaatcgcgccacacctggaactgtgtgtcctctggctcag 720

DB 3497 AAGCCTACGAGACGACCGAATCCGCCACCCCTGAGCTGTGTCTCTGGGCTCAG 3438  
 QY 721 gggaaaggtatgcccacaacacacagacggtggccgttgactgggtcctacatcgtgc 780  
 DB 3437 GGGAAAGGGTGTGCTCCAAACACCAAGACCGTGGCGGTGACTGGGCTCATCTCGGTC 3378  
 QY 781 cccccgaagtggaggttccaaactccggcccttataccctattctctagtaact 840  
 DB 3377 CTCCCGAGGTGAGCGCTTCAAACTCCGGCCCTTATACCTCATCTGATGATCAT 3318  
 QY 841 ggcacatcctagatgacggtcgttcaagaagcttctatcagaggtctgtgtggtgtg 900  
 DB 3317 GCGATCTCTTATGATGACCGTGTTCACAGCGTTTATGACAGGCTGTGTGGCTGTG 3258  
 QY 901 gccaggggtgaactcccgcgcaatatctattgctcgtcctaagtcggtatcagat 960  
 DB 3257 GCCAGGGGTGAACCTCCCGCAATATCTATTGCTCTTATGCGGGGTATCCGAGAT 3198  
 QY 961 tacacgagatggccatctgacacgacgagagcttacctagtagaagacccaagaagggcc 1020  
 DB 3197 TACAGGATGGCCATCTGTACACGACCGGAGGCTACTAGTAGAAGCCTCAAGAGAGCC 3138  
 QY 1021 tcctga 1026  
 DB 3137 TCCTGA 3132

RESULT 3  
 AAQ37036  
 ID AAQ37036 standard; DNA; 615 BP.  
 XX  
 AC AAQ37036;  
 XX  
 DT 05-JUL-1993 (first entry)  
 XX  
 DE SOD-T gene.  
 XX  
 KW Super oxide dismutase; heat resistant; medicine; inflammation;  
 KW X-ray; enzyme electrodes; bioreactor; ss.  
 OS  
 XX *Thermus aquaticus*.  
 PN JP05030976-A.  
 XX  
 XX 09-FEB-1993.  
 PD  
 XX 26-JUL-1991; 91JP-0208747.  
 PF  
 XX 26-JUL-1991; 91JP-0208747.  
 PR  
 XX (YONS-) YONSUBA NYUGYO KK.  
 PA  
 XX WPI: 1993-087958/11.  
 DR P-PDB: AAR32376.  
 DR  
 XX

PT Gene, of and plasmid contg. heat resistant super:oxide dismutase  
 PT - for treating inflammation, X-ray protection, enzyme electrodes  
 PT and bio:reactor enzymes  
 PS  
 XX Claim 1: Page 7; 9pp; Japanese.  
 XX  
 CC The sequence shown is that isolated from *Thermus aquaticus* genomic  
 CC DNA and encodes a heat resistant super oxide dismutase enzyme. The  
 CC gene is useful for medicines for inflammation and as a protection  
 CC against X-rays, as a reagent for research in enzyme electrodes or as  
 CC an enzyme for a bioreactor.  
 XX

Sequence 615 BP; 125 A; 223 C; 183 G; 84 T; 0 other;

Query Match 5.3%; Score 54.4; DB 14; Length 615;  
 Best Local Similarity 47.5%; Pred. No. 0.0023;

PS CLAIM 3; Page 60-64; 101pp; English

```

ET      /trans_except= (pos: 1::2, aa: 1ML-aaad-aaad-aaad-aaad,
        /product= "ORF#1 protein"

```

```

FT      /note= "encoded protein shown in AAW72095, Xaa=
FT      unspecified amino acid"
FT      complement (5962..9306)
FT      /tag= b
FT      /product= "ORF#2 protein"
FT      /note= "encoded protein shown in AAW72096"
FT      9246..11195
FT      /tag= c
FT      /product= "ORF#3 protein"
FT      /note= "encoded protein shown in AAW72097"
FT      11657..15199
FT      /tag= d
FT      /product= "ORF#4 protein"
FT      /note= "encoded protein shown in AAW72098"
FT      15252..16265
FT      /tag= e
FT      /product= "ORF#5 protein"
FT      /note= "encoded protein shown in AAW72099"
FT      complement (16548..17621)
FT      /tag= f
FT      /product= "ORF#6 protein"
FT      /note= "encoded protein shown in AAW72100"
FT      18521..19936
FT      /tag= g
FT      /product= "ORF#7 protein"
FT      /note= "encoded protein shown in AAW72101"
FT      20015..21034
FT      /tag= h
FT      /transl_except= (pos: 21033..21034, aa: Gly-Ser)
FT      /product= "ORF#8 protein"
FT      /note= "encoded protein shown in AAW72102"
FT      W09820016-A1.
FT      14-MAY-1998.
FT      31-OCT-1997; 97WO-US20016.
FT      09-JUN-1997; 97US-0049018.
FT      04-NOV-1996; 96US-0030279.
FT      (SMK ) SMITHKLINE BEECHAM CORP.
FT      Chan JY, Dabrowski-Amara CE, Delvecchio AM, Dillon SB;
FT      Esser KM, Leary JJ;
FT      WPI: 1998-286847/25.
FT      P-PSDB: AAW72095, AAW72096, AAW72097, AAW72098, AAW72099, AAW72100,
FT      AAW72101, AAW72102.
FT      Herpes simplex virus type-2 sequences - useful in, e.g. prevention
FT      and treatment of infection or inducing immunological response in
FT      mammal
FT      Claim 1: Page 350-359; 74BPP; English.
FT      This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA
FT      sequence of the invention. This sequence was isolated from HSV-2 strain
FT      SB5 (deposited as ATCC VR-2546), is designated Contig ID 10, and encodes
FT      8 HSV-2 proteins. The proteins can be used for the treatment or
FT      prevention of disease, to induce an immunological response in a mammal or
FT      to identify inhibitors, activators or novel antivirals. Antagonists of
FT      the proteins can be used to inhibit a viral polypeptide. The DNA sequence
FT      or a vector containing it can also be used to induce an immunological
FT      response in a mammal.
FT      Sequence 21034 BP; 3154 A; 7419 C; 7234 G; 3227 T; 0 other;

```

Query Match 4.7%; Score 48.6; DB 19; Length 21034;  
 Best Local Similarity 43.9%; Pred. No. 0.12;  
 Matches 207; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

```

QY      355  ggcacccggagctctgaggaagtgggagcgtcttcttattgttccctggagatg 414
QY      1211  |||||
Db      2012  gccatgacagagctcagacccttggtgggtcgccgcccgcgacctggtatccagggccgc 1953
QY      415  ttggccctcaacctgggggttcacccgcagaccgtccacgcctggaagaagttcctttag 474
QY      1111  |||||
Db      1952  cgggcttgccgcccacgcccaggggacacatgcccctcattgatgatcccgccgcccgcgc 1893
QY      475  aaaaaggccctgtgtgcacacgagcttccttcacccaacgcgtcaacgaggagcgccggcc 534
QY      1111  |||||
Db      1892  acgaccgctgcttcgggaaagccctcgccggggcagacagagcttggttcgggggctgtgcac 1833
QY      535  atcgagacccttggcgcgttcgcgtgaggaagcaggaagcagagctacacctggagcagc 594
QY      1832  gccgaaggacggccggggacatttccccacggggcgcccttgagagctgggccaag 1773
QY      595  tacatctaccccttgaggaacctcgccctagacatgtgcacagcgctgtcttcctcaac 654
QY      1111  |||||
Db      1772  gmtatggcgccacgcggcagccgcccgcagacaaattgagggccgcttcgcccagctgacg 1713
QY      655  ttggtcaaggcctaccaggaacacgaatccgcccacccttgagcgtgtgtcctctgg 714
QY      1712  gccgaagatggcgccgcccagccgcccgcagacgacgctgggcccgcgctgcag 1653
QY      715  gctcaagggaagaaaggltgattgcacacacaaagacgctgacgttgcctgagcctcacc 774
QY      1111  |||||
Db      1652  cgggcttgacccgctgacccctcagaaacccgcccgaatttcagctgagctgacgctgcgc 1593
QY      775  ctggtctcccgaggttgagcgttcacaaatcccgccctatcacccctc 825
QY      1111  |||||
Db      1592  caggcctgagggcgccagcagcagcgtacacccccgggacttccgcaagcgc 1542

RESULT 6
AAV62176/c
ID      AAV62176 standard; DNA; 117213 BP.
XX      AC
XX      AAV62176;
XX      13-JAN-1999 (first entry)
DT      DE
XX      XX
XX      HSV-2 strain SB5 Contig ID 15 DNA sequence.
XX      XX
XX      HSV-2 strain SB5; immunological response induction; therapy;
XX      KM      antiviral identification; viral protein inhibitor; ss.
XX      OS      Herpes simplex virus type 2.
XX      XX
XX      Key      Location/Qualifiers
FH      755..1297
FT      CDS
FT      /tag= a
FT      /product= "ORF#1 protein"
FT      /note= "encoded protein shown in AAW72170"
FT      1170..2174
FT      /tag= b
FT      /product= "ORF#2 protein"
FT      /note= "encoded protein shown in AAW72171"
FT      2229..2930
FT      /tag= c
FT      /product= "ORF#3 protein"
FT      /note= "encoded protein shown in AAW72172"
FT      complement (3130..3735)
FT      /tag= d
FT      /product= "ORF#4 protein"
FT      /note= "encoded protein shown in AAW72173"
FT      complement (3802..6447)
FT      /tag= e
FT      /product= "ORF#5 protein"
FT      /note= "encoded protein shown in AAW72174"
FT      6017..8482
FT      /tag= f
FT      /product= "ORF#6g protein"
FT      /note= "encoded protein shown in AAW72250"
FT      CDS

```

FT CDS 6026..8482  
FT /\*tag- g  
FT /product- "ORF#6f protein"  
FT /note- "encoded protein shown in AAW72249"  
FT 6065..8482  
FT /\*tag- h  
FT /product- "ORF#6e protein"  
FT /note- "encoded protein shown in AAW72248"  
FT 6167..8482  
FT /\*tag- i  
FT /product- "ORF#6d protein"  
FT /note- "encoded protein shown in AAW72247"  
FT 6296..8482  
FT /\*tag- j  
FT /product- "ORF#6c protein"  
FT /note- "encoded protein shown in AAW72246"  
FT 6326..8482  
FT /\*tag- k  
FT /product- "ORF#6b protein"  
FT /note- "encoded protein shown in AAW72245"  
FT 6446..8482  
FT /\*tag- l  
FT /product- "ORF#6a protein"  
FT /transl\_except- (pos: 7400..7402, aa: Ala-Ala)  
FT /transl\_except- (pos: 7481..7486, aa: Ile)  
FT /note- "encoded protein shown in AAW72244"  
FT 8457..9347  
FT /\*tag- m  
FT /product- "ORF#7 protein"  
FT /note- "encoded protein shown in AAW72175"  
FT complement (9604..11855)  
FT /\*tag- n  
FT /transl\_except- (pos: 11635..11636, aa: Ala)  
FT /product- "ORF#8 protein"  
FT /note- "encoded protein shown in AAW72176"  
FT complement (11905..14508)  
FT /\*tag- o  
FT /product- "ORF#9b protein"  
FT /note- "encoded protein shown in AAW72222"  
FT complement (11905..14520)  
FT /\*tag- p  
FT /product- "ORF#9a protein"  
FT /note- "encoded protein shown in AAW72223"  
FT 14399..15802  
FT /\*tag- q  
FT /product- "ORF#10 protein"  
FT /note- "encoded protein shown in AAW72177"  
FT complement (15996..16286)  
FT /\*tag- r  
FT /product- "ORF#11 protein"  
FT /note- "encoded protein shown in AAW72178"  
FT complement (16202..18064)  
FT /\*tag- s  
FT /product- "ORF#12 protein"  
FT /note- "encoded protein shown in AAW72179"  
FT complement (18105..19661)  
FT /\*tag- t  
FT /product- "ORF#13 protein"  
FT /note- "encoded protein shown in AAW72180"  
FT complement (19415..20074)  
FT /\*tag- u  
FT /product- "ORF#14 protein"  
FT /note- "encoded protein shown in AAW72181"  
FT 20155..21453  
FT /\*tag- v  
FT /product- "ORF#15 protein"  
FT /note- "encoded protein shown in AAW72182"  
FT complement (21326..22291)  
FT /\*tag- w  
FT /product- "ORF#16 protein"  
FT /note- "encoded protein shown in AAW72183"  
FT complement (22546..24654)  
FT /\*tag- x  
FT CDS

FT /\*product- "ORF#17 protein"  
FT /note- "encoded protein shown in AAW72184"  
FT 24684..25955  
FT /\*tag- y  
FT /product- "ORF#18 protein"  
FT /note- "encoded protein shown in AAW72185"  
FT complement (26295..27251)  
FT /\*tag- z  
FT /product- "ORF#19 protein"  
FT /note- "encoded protein shown in AAW72186"  
FT complement (27630..31754)  
FT /\*tag- aa  
FT /product- "ORF#20b protein"  
FT /note- "encoded protein shown in AAW72225"  
FT complement (27630..31784)  
FT /\*tag- ab  
FT /product- "ORF#20a protein"  
FT /note- "encoded protein shown in AAW72224"  
FT complement (32067..32735)  
FT /\*tag- ac  
FT /product- "ORF#21 protein"  
FT /note- "encoded protein shown in AAW72187"  
FT 33140..34984  
FT /\*tag- ad  
FT /product- "ORF#22a protein"  
FT /note- "encoded protein shown in AAW72226"  
FT 33386..34984  
FT /\*tag- ae  
FT /product- "ORF#22b protein"  
FT /note- "encoded protein shown in AAW72227"  
FT complement (35205..37721)  
FT /\*tag- af  
FT /product- "ORF#23 protein"  
FT /note- "encoded protein shown in AAW72188"  
FT complement (38058..39188)  
FT /\*tag- ag  
FT /product- "ORF#24 protein"  
FT /note- "encoded protein shown in AAW72189"  
FT 39090..39935  
FT /\*tag- ah  
FT /product- "ORF#25 protein"  
FT /note- "encoded protein shown in AAW72190"  
FT 40216..41973  
FT /\*tag- ai  
FT /product- "ORF#26 protein"  
FT /note- "encoded protein shown in AAW72191"  
FT 42206..44178  
FT /\*tag- aj  
FT /transl\_except- (pos: 44063..44064, aa: Lys)  
FT /product- "ORF#27 protein"  
FT /note- "encoded protein shown in AAW72192"  
FT complement (44853..47297)  
FT /\*tag- ak  
FT /product- "ORF#28 protein"  
FT /note- "encoded protein shown in AAW72193"  
FT 47122..47338  
FT /\*tag- al  
FT /product- "ORF#29 protein"  
FT /note- "encoded protein shown in AAW72194"  
FT complement (47305..49662)  
FT /\*tag- am  
FT /product- "ORF#30 protein"  
FT /note- "encoded protein shown in AAW72195"  
FT complement (50035..51666)  
FT /\*tag- an  
FT /product- "ORF#31 protein"  
FT /note- "encoded protein shown in AAW72196"  
FT complement (51701..53575)  
FT /\*tag- ao  
FT /product- "ORF#32 protein"  
FT /note- "encoded protein shown in AAW72197"  
FT 54393..58115  
FT /\*tag- ap  
FT CDS

```

FT FT /product= "ORF#33 protein"
FT FT /note= "encoded protein shown in AAW72198"
FT FT complement (58060..58977)
FT FT /tag= ag
FT FT /product= "ORF#34 protein"
FT FT /note= "encoded protein shown in AAW72199"
FT FT complement (58970..60760)
FT FT /tag= ar
FT FT /product= "ORF#35 protein"
FT FT /note= "encoded protein shown in AAW72200"
FT FT 60759..61151
FT FT /tag= as
FT FT /product= "ORF#36 protein"
FT FT /note= "encoded protein shown in AAW72201"
FT FT 61241..62071
FT FT /tag= at

```

Query Match 4.7%; Score 48.6; DB 19; Length 117213;  
 Best Local Similarity 43.9%; Pred. No. 0.19;  
 Matches 207; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

```

355 ggcacccggagctcttgaggaagtgaggagctgtcttcaatgctccctggagatg 414
    ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68405 GCCATGACGACGCTCAGACACCTGGGTGGCGGCCGCCGACCTGTATGTCAGGCGCGC
QY 415 ttggcctcaacttggtggtacccggacagacgttccacgcttgaagaagtgcttga 474
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68345 CGGGCTGCGGCCGACGCGGAGGGGACCATGCGCCCTCATCATGATGCGGCGCGCGC
QY 475 aaaaaggagctgtgagccacgagcttcttcaacaacgctcaacggagagcgcgagcc 534
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68285 ACGACCCCTGCTGGGAAAGCTTCGGCGGGACAGGCTGTGGGGGGGCTGTGAC 68226
QY 535 atcggaccccttgggagctgctgagagcagaggaagcagagctcaccctggagcag 594
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68225 GCGGAGGAGCGCGCGGACCATTCCTCCAGGGGGCGCCCTGACAGAGCTGGCGAAG 68166
QY 595 tacatcacccttgaggaacccctgacccctgaatgagcagggagtgctcttcaac 654
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68165 GTGATCGGCGCCACGCGACCCCGCGGACGAACTTGAGGCCCGCTCGCGGACCTGACG 68106
QY 655 ttggtaaggcttaccaggaacacggaatccgcccacccttgagctgtgtctcttgg 714
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68105 GCGAAGATGGCGGCCGACGCGGCCCGGGGAGAGGAGCGAGCGCTGGGCGCGCTGAG 68046
QY 715 gtcacgggaaaaggtgatgtcccaacacgaagccgtgagcgttgaacctgggctcacc 774
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68045 GCGGCCCTGGACCGGCTGACGAAACCGCGCGGAGTTGACGTGGTGAAGCTGGCGCGCTG 67986
QY 775 ctgtctctcccgaggtgagagcttccaaactcccgacctataacctc 825
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 67985 CAGGCGCTGGCGGACGACGAGGCTAACACCCCGGACTTCCGCAAGCGC 67935

```

RESULT 7  
 AAD25519/c  
 ID AAD25519 standard; DNA: 154746 BP.  
 XX AAD25519;  
 AC AAD25519;  
 DT 26-MAR-2002 (first entry)  
 DE Human herpesvirus 2 complete DNA genome.  
 XX  
 KW Human herpesvirus 2; cytosolic; cancer; immunosuppressive; virucide;  
 KW antibacterial; fungal; protozoicide; antirheumatic; antiinflammatory;  
 KW antiaerobic; rheumatoid arthritis; neuroprotective; multiple sclerosis;  
 KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;  
 KW vasculitis; ds.  
 OS Human herpesvirus 2.  
 XX  
 PN WO200176643-A1.

```

XX 18-OCT-2001.
PD 06-APR-2001; 2001WO-US11372.
XX 07-APR-2000; 2000US-195680P.
XX (BAYU) BAYLOR COLLEGE MEDICINE.
XX Orson FM, Kinsey BM, Bhogal BS;
XX WPI; 2002-066308/09.
XX
XX Composition for oral delivery of vaccines, comprises expression vector
XX containing antigenic genomic sequence, bound to aggregated
XX protein-polycationic polymer conjugate or suspension
XX
XX Disclosure: Page 90-132; 145pp; English.
XX
XX The invention relates to a composition comprising an expression vector
XX bound to an aggregated protein-polycationic polymer conjugate or
XX suspension. The expression vector contains a promoter polynucleotide
XX sequence operatively linked to a polynucleotide sequence encoding an
XX antigen which is a fragment of a gene or genome associated with an
XX infectious disease, cancer and autoimmune disease such as rheumatoid
XX arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
XX consisting of bacterium, fungus, protozoa and virus such as human
XX immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
XX virus (HCV), influenza and respiratory syncytial virus (RSV), and
XX optionally comprising a nucleotide sequence encoding a cytokine (or a
XX cytokine expression vector), is useful for inducing an immune response
XX (systemic and/or mucosal) in an organism. The cytokine expression vector
XX contains a sequence for granulocyte macrophage-colony stimulating factor
XX (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
XX the antigen and the cytokine are under transcriptional control of same or
XX different promoter polynucleotide sequences. The expression vector, as a
XX DNA vaccine is useful for treating a condition in an organism. The
XX present sequence is human herpesvirus 2 complete DNA genome related
XX to the invention.
XX
XX Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 other;

```

Query Match 4.7%; Score 48.6; DB 24; Length 154746;  
 Best Local Similarity 43.9%; Pred. No. 0.2;  
 Matches 207; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

```

QY 355 ggcacccggagctcttgaggaagtgaggagctgtcttcaatgctccctggagatg 414
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 77292 GCCATGACGACGCTCAGACACCTGGGTGGCGGCCGCCGACCTGTATGTCAGGCGCGC 77233
QY 415 ttggcctcaacttggtggtacccggacagacgttccacgcttgaagaagtgcttga 474
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 77232 CGGGCTGCGGCCGACGCGGAGGGGACCATTCCTCATCATGATGCGGCGCGCGCGC 77173
QY 475 aaaaaggagctgtgagccacgagcttcttcaacaacccgtcaacggagagcgcgagcc 534
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 77172 ACGACCGCTGCTGGGAAAGCTTCGGCGGGACAGAGCTGTGGGGGCTGTCTCAC 77113
QY 535 atcggaccccttgggagctgctgagcaggaagcagacgctcacccttgaagcag 594
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 77112 GCGAAGGAGCGCGCGGACCATTCCTCCCAAGCGCGGCCCTGACGAGCTGGGCAAG 77053
QY 595 tacatcacccttgaggaacccctgacccctgaatgacagggcggtcttcttcaac 654
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 77052 GTCATGCGCGCACGACGCGCGCGGACGAACTTGAAGCCGCGCGGACCTGACG 76993
QY 655 ttggtaaggcttaccaggaacacggaatccgcccacccttgagagctgtgtctctgg 714
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 76992 GCGAAGATGGCGCGCGCGCGCGGAGGACGAGCGCTGTGGCGCGCGCGCTGAG 76933
QY 715 gtcacgggaaaaggtgatgtcccaacacgaagccgttgaacctgggctcacc 774
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

CC used in the method of the invention.  
XY

PR	30-MAY-1997;	97US-0048189.
22	20-MAY-1997:	97US-0048357.

PR 30-MAY-1997; 97US-0048100.

PR 30-MAY-1997; 97US-0050934.  
PR 06-JUN-1997; 97US-0048970.  
PR 05-SEP-1997; 97US-005765.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ferrie AM, Fischer CL, Gentz RL, Greene JM, Kyaw H;  
PI Li H, Li Y, Moore PA, Rosen CA, Ruben SM, Soppet DR;  
PI Wei YF, Young PE, Zeng Z;

XX WPI: 1998-520811/44.  
DR P-PSDB: AAW75227.

PT Isolated human poly(nucleotide(s) encoding secretory peptide(s) -  
PT used to develop products for the diagnosis and treatment of e.g.  
PT inflammation, cancers, CNS disorders or immune system disorders

PS Claim 1; Page 144-145; 201pp; English.

XX This sequence represents a nucleic acid molecule which encodes a  
XX secreted human protein. The gene number, and the clone it is derived  
XX from, are detailed in the descriptor line. The gene can be used to  
XX generate fusion proteins by linking to the gene to a human immunoglobulin  
XX Fc portion (e.g. AAV34277) for increasing the stability of the fused  
XX protein as compared to the human protein only.  
XX The invention relates to 28 novel genes and their fragments (nucleic  
XX acid sequences: AAV34286-V34325; amino acid sequences AAW75196-W75235)  
XX which are useful for preventing, treating or ameliorating medical  
XX conditions e.g. by protein or gene therapy. Also, pathological  
XX conditions can be diagnosed by determining the amount of the new  
XX polypeptides in a sample or by determining the presence of mutations in  
XX the new polynucleotides. Specific uses are described for each of the 28  
XX polynucleotides, based on which tissues they are most highly expressed in  
XX (see AAV34286 for described uses).

SQ Sequence 1473 BP; 257 A; 476 C; 491 G; 249 T; 0 other;

Query Match 4.5%; Score 46.2; DB 19; Length 1473;

Best Local Similarity 42.5%; Pred. No. 0.23;

Matches 299; Conservative 0; Mismatches 403; Indels 1; Gaps 1;

```

OY 314 tgaactctccagagattgcccgtccacatccctcctgaagcaggagcttga 373
DB 823 TGGAACTGCTCTGACTATTCCAAATCTCCAGGCCACACAGCAGCAAGTTGAGT 764
OY 374 ggaagctgggagcgtgcttcatgtcccccctgagagatgttgccccaacctggg 433
DB 763 TTCTGATATAGCTTGTCCAGTGATTCACCACTCCAGCACCAGATGACGACGGCTGCT 704
OY 434 taaccggagacgtccacgctggaagaagctccttgagaanaaaggcctgtgtgcca 493
DB 703 TCGCTCCCGGATCTGGAATCTGCATTTGTCACGACCAACATACAGCTCCAGTACGCG 644
OY 494 ccgagctccttcaacaacgctcaaggagagcgccggcctatggc-acctttgggc 552
DB 643 GTCCTCCGGGATGGAGAAAGTCCCGGGCCAGGCTTAAGAGGCTCCGTCGGGGT 584
OY 553 gtccggctgagccagaggaagcagctacccctgagactatactaccctggag 612
DB 583 CCCAGAGGCTCCAGAGGCTGTGTCGCTGAGACCCGACAGTCCCGGCTGTGACGACAG 524
OY 613 aacctgccttgaatgagcagcggtgtctctccttcaactgggtgaagcctaccag 672
DB 523 TCGTCAAGCTGTATACAGGCGTCCGCTCCGCTCCCACTTCATCCAGGGGCTCGATC 464
OY 673 gaccacgaatccgccacccttgagcgtgtgtcctctggtcctcaggggaaggtg 732
DB 463 AGGTCAAGTCTGACCCACCTGGAAGAACCCCTGAGCGCGGACAGAGTGTGAGCGTG 404
OY 733 atgccacaacacgaagcgttgccgttgacctgtgacctcactcctcccgaggtg 792
DB 403 GCGGCTGAGTCCGTTACCTCTTAAGTGGCCCTGTGTAAGACAGAGTGTGCTCCGCGAGG 344

```

```

OY 793 gaggcttccaaactccggccctatataccctcattgtctaagtaattgccgcatccta 852
DB 343 CTGCTCCGTCACCTCGGAGGCGATTTGGCAGCGGATAGTCTCTGTGTAGCCGAGCCAG 284
OY 853 gatgaccgtgcttcaagaagcttctatcagaggttctgtggtgctgtgccaagggtgaa 912
DB 283 CAGGTCCCTGTTCTTCGCGAGGTGGAGGTGAAGTTGTGCCCTGTGGCCCCCAAGAGAGTA 224
OY 913 ctcccgcgcaatatctatttgccttactatgctgaggttaccagattacaagatgac 972
DB 223 GCTCACCTCTCTGTGGGTGCGAGGCCCAAGTGGAGGGGAGAGCTCGCGGACTCGGGGCC 164
OY 973 catctgacacgacggagagcgtactagtgaaagacctcaag 1015
DB 163 TGGCAGACGCCACGCGCAACACAGACCTCATCTGCTCATGAGG 121

```

RESULT 10

AAQ22482  
ID AAQ22482 standard; DNA; 1620 BP.

XX AAQ22482;

AC 29-JUL-1992 (first entry)

DE groEL-1 gene coding region.

XX Heat shock protein; HSP18; ss.

OS Streptomyces albus.

XX Key Location/Qualifiers

FT CDS 1..510 /tag= a /note= "groEL-1 gene encoding HSP18"

PN W09204452-A.

PD 19-MAR-1992.

PF 03-SEP-1991; 91WO-FR00701.

PR 10-SEP-1990; 90FR-0011186.

PA (INSP) INST PASTEUR.

PI Mazodier P, Guglielmi G;

DR WPI: 1992-114358/14.

DR P-PSDB: AAR22363.

PT Recombinant DNA contg. heat inducible promoter and heterologous

PT gene - also vectors, transformed cells and new heat shock

PT proteins of Streptococcus albus

PS Claim 18; Fig 8; 50pp; French.

XX The sequence is that of the coding region of the groEL-1 gene which

XX codes for the 18 kDa heat shock protein HSP18. The gene actually

XX CC codes for a 56 kDa protein but this is subjected to post translation

XX CC modification to give the 18 kDa HSP18 protein.

XX See also AAQ22477-Q22486.

SQ Sequence 1620 BP; 287 A; 590 C; 539 G; 204 T; 0 other;

Query Match 4.4%; Score 45.6; DB 13; Length 1620;

Best Local Similarity 43.1%; Pred. No. 0.33;

Matches 219; Conservative 0; Mismatches 289; Indels 0; Gaps 0;

```

OY 289 gtcaagagcggcctcagaagctcttgaaactcctcagaagatgccgctcacatc 348

```

Db 1081 gaccgagagaagctccagagcgctcgcacagctggccgagcgctcgtatccgc 1140  
 QY 349 cccatcgcaacccgagctctgagagaaggctgggagcgctcctcatcgtccctc 408  
 Db 1141 gtcggcgagccacccagagctgagctgagagagcgacacgctctggagagccatc 1200  
 QY 409 gagatgttgccctcaactctgggggtcaaccgagacgctccacgctggaaagatc 468  
 Db 1201 tccgagaccgcgcgcgcgtctgagagagagcatcgtctcggctgggtccgcgtggtc 1260  
 QY 469 cttagagaaaggcgctctgtgacacgctccctcaacacgtcaacggggagcg 528  
 Db 1261 cagcgctcaagagctcctggagacacactcggccgacggcgagagccacgggtgtc 1320  
 QY 529 cgggacatcgacacactcttgagcgctccggtgagcgagagaaagccacgctc 588  
 Db 1321 ggggtcgtccgagcgccgctcgagcgctgctgagctgacgcgagaaagcgccctc 1380  
 QY 589 gacgactacatcaccctctggagagaaactcgcctcagacagcgccagcggtgtc 648  
 Db 1381 gagggctcgtatcatcaacacaaaggtggcgagctcgacaaaggccagggctcaacg 1440  
 QY 649 ttcaactgggtcaagagctcctacagacagacggaatccgcccacacttgagctgtc 708  
 Db 1441 gccacccggagagtaagcgacactgtgtcaagcgcgctcatcgacccggtcaaggtcac 1500  
 QY 709 cctcgggtcagggagaaagggtgatgccaacacaaagacggtggcgttgaactgggc 768  
 Db 1501 cgcctcgccctggagaaagcgagcctccatcgctcctcctgtccttgagacgagacctg 1560  
 QY 769 ctcatcctgtctcccgaggtgagc 796  
 Db 1561 gtcgtcgaagaagcgcgccgagagagagc 1588

## RESULT 11

AAQ22485

ID AAQ22485 standard: DNA; 2668 BP.

XX AC AAQ22485;

XX DT 29-JUL-1992 (first entry)

XX DE groEL-1 gene.

XX KW Heat shock protein; HSP18; ss.

XX ST Streptomyces albus.

XX KEY Location/Qualifiers

XX CDS 849..2468 /tag= a /note= "Gro EL1 protein"

XX FT

XX XX WO9204452-A.

XX XX 19-MAR-1992.

XX PF 03-SEP-1991; 91MO-FR00701.

XX PR 10-SEP-1990; 90FR-0011186.

XX PA (INSP ) INST PASTEUR.

XX PI Mazodier P, Guglielmi G;

XX XX WPI; 1992-114358/14.

XX DR P-PSDB; AAR22363.

XX PT Recombinant DNA contg. heat inducible promoter and heterologous  
 XX gene - also vectors, transformed cells and new heat shock  
 XX proteins of Streptococcus albus

PS Claim 18; Fig 8; 50pp; French.

XX CC The sequence is that of the complete groEL-1 gene which codes for  
 XX CC the 18 kDa heat shock protein HSP18. The gene actually codes for a  
 XX CC 56 kDa protein but this is subjected to post translation modification  
 XX CC to give the 18 kDa HSP18 protein. See also AAQ22477-Q22486.

XX SQ Sequence 2668 BP; 443 A; 948 C; 929 G; 348 T; 0 other;

Query Match 4.48; Score 45.6; DB 13; Length 2668;

Best Local Similarity 43.18; Pred. No. 0.37; Mismatches 289; Indels 0; Gaps 0;

Matches 219; Conservative

QY 289 gtcagagagcggtcctgaagctcttgaaactcctccacagagatggccgctccacatc 348  
 Db 1929 gaccgagagaagctcccgagagcgctcgcacagctggcgcggtcgtcgtatccgc 1988  
 QY 349 cccatcgcaacccgagctctgagagaaggctgggagcgctcctcatcgtccctc 408  
 Db 1989 gtcggcgagccacccagagctcgaagctgaagagcgacacgctctggagagccatc 2048  
 QY 409 gagatgttgccctcaactctgggggtcaaccgagacggctccacgcttgaagatc 468  
 Db 2049 tccgagaccgcgcgcgcgtctgagagagagcatcgtctcgggtgtgctccgcgtgtc 2108  
 QY 469 cttagagaaaggcgctctgtgacacgagctcctcaacacgtcaacggggagagc 528  
 Db 2109 cagcgctcaagagctcctggagacacactcggccgacagcgagagccacgggtgtc 2168  
 QY 529 cgggacatcgacacactcttgagcgctccggtgagggccagagaaagccaggtcacctg 588  
 Db 2169 ggggtcgtccgagcgccgctcgaagcgctgagctgtgagctgcgcgagaaagcgcgctc 2228  
 QY 589 gacgactacatcaccctctggagagaaactcgcctcagacagcgccaaagcggtgtc 648  
 Db 2229 gagggctcgtatcatcaacacaaaggtggagctcgacaaaggcgacgttcaacg 2288  
 QY 649 ttcaactgggtcaagagctcctacagacagacggaatccgcccacacttgagctgtc 708  
 Db 2289 gccacccggagagtaagcgacactgtgtcaagcgcgctcatcgacccggtcaaggtcac 2348  
 QY 709 cctcgggtcagggagaaagggtgatgccaacacaaagacggtggcgttgaactgggc 768  
 Db 2349 cgcctcgccctggagaaagcgagcctccatcgctcctcctccttgagacgagacctg 2408  
 QY 769 ctcatcctgtctcccgaggtgagc 796  
 Db 2409 gtcgtcgaagaagcgcgccgagagagagc 2436

## RESULT 12

AAS54366

ID AAS54366 standard: DNA; 1686 BP.

XX AC AAS54366;

XX DT 13-FEB-2002 (first entry)

XX DE Pseudomonas aeruginosa DNA for cellular proliferation protein #497.

XX KW Antisense; ds; prokaryotic cellular proliferation gene;

XX KW antibiotic; antibacterial; drug design.

XX OS Pseudomonas aeruginosa.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001MO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.



PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GI;  
PI Yamamoto RT, Xu HH;  
XX  
DR WPI: 2001-611495/70.  
DR P-PSDB; AAU36507.  
XX  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
PS Claim 27; Seq ID No 8003; 511pp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
XX Sequence 1686 BP; 307 A; 607 C; 528 G; 244 T; 0 other;

XX  
DT 07-DEC-1998 (first entry)  
XX  
XX HSV-2 strain SBS Contig ID 104 DNA sequence.  
DE  
XX HSV-2 strain SBS; immunological response induction; therapy;  
KW antiviral identification; viral protein inhibitor; ss.  
KW  
XX Herpes simplex virus type 2.  
OS  
XX  
FH Location/Qualifiers  
XX Key  
XX CDS  
XX  
XX /tag= a  
XX /product= "ORF#1 protein"  
XX /note= "encoded protein shown in AAW72033"  
XX complement (2895..3908)  
XX  
XX /tag= b  
XX /product= "ORF#2 protein"  
XX /note= "encoded protein shown in AAW72034"  
XX complement (3637..6093)  
XX  
XX /tag= c  
XX /product= "ORF#3 protein"  
XX /note= "encoded protein shown in AAW72035"  
XX complement (6263..6745)  
XX  
XX /tag= d  
XX /product= "ORF#4 protein"  
XX /note= "encoded protein shown in AAW72036"  
XX complement (6775..8046)  
XX  
XX /tag= e  
XX /product= "ORF#5 protein"  
XX /note= "encoded protein shown in AAW72037"  
XX complement (8386..9477)  
XX  
XX /tag= f  
XX /product= "ORF#6 protein"  
XX /note= "encoded protein shown in AAW72038"  
XX complement (9721..13911)  
XX  
XX /tag= g  
XX /product= "ORF#7 protein"  
XX /note= "encoded protein shown in AAW72039"  
XX complement (14158..14826)  
XX  
XX /tag= h  
XX /product= "ORF#8 protein"  
XX /note= "encoded protein shown in AAW72040"  
XX complement (15162..17075)  
XX  
XX /tag= i  
XX /product= "ORF#9 protein"  
XX /note= "encoded protein shown in AAW72041"  
XX complement (17295..19110)  
XX  
XX /tag= j  
XX /product= "ORF#10 protein"  
XX /note= "encoded protein shown in AAW72042"  
XX complement (19111..20064)  
XX  
XX /tag= k  
XX /product= "ORF#11 protein"  
XX /note= "encoded protein shown in AAW72043"  
XX complement (20149..21279)  
XX  
XX /tag= l  
XX /product= "ORF#12 protein"  
XX /note= "encoded protein shown in AAW72044"  
XX complement (20962..22026)  
XX  
XX /tag= m  
XX /product= "ORF#13 protein"  
XX /note= "encoded protein shown in AAW72045"  
XX complement (22307..23887)  
XX  
XX /tag= n  
XX /product= "ORF#14 protein"  
XX /note= "encoded protein shown in AAW72046"  
XX complement (23778..24065)  
XX  
XX /tag= o  
XX /product= "ORF#15 protein"  
XX /note= "encoded protein shown in AAW72047"  
XX complement (24286..25632)  
XX  
XX /tag= p



PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140981.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145277.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151067.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154479.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 13-OCT-1999; 99US-0159325.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159367.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 4.4%; Score 45.2; DB 21; Length 1290;  
Best Local Similarity 47.2%; Pred. No. 0.38;  
Matches 137; Conservative 0; Mismatches 153; Indels 0; Gaps 0;  
Qy 479 agggcctgtgtggcaccgacgtcttcacccaacgltcaacgggagcgccggccatcg 538  
Db 76 aagttcaagcgcggcaccattgatgtaacccgccaagatcgcgcgccgctccg 135

QY	539	gacaccttggcgccgtccgctgtgagggccaggaaagccacgagctcaccccttgagcagctacta	538
QY	539	gacaccttggcgccgtccgctgtgagggccaggaaagccacgagctcaccccttgagcagctacta	538
Db	136	cctgcgtccgctgcgtccgcccgtctgtgcgaactctgcacaggcgcttcgcgtcgcagccg	195
QY	539	ctacaccttgaggaaggaactcgcctctagacatgctgcacagcgctgcctctccttaactgg	658
Db	196	ggcgccgacagatcacgctgcctcctccatgcttcgcgcgtcaggaagcttcgcacgaagtgcg	255
QY	659	tcaagcgctaccagagaccacgaaatccgcgccacccttgagcgtgctgtctccttgagctc	718
Db	256	tgcagcgccgacgaagctgcgcgtcttcgcgtcgcgcacactcgcgcgctcctcctcgtcggcg	315
QY	729	aggggaaaagggtgatgctccacacccaagaccgtgagccgttgacactggcgc	768
Db	316	cgaagcgaggaggcacgcgcccaagagctgtacctacgacagatccagac	365
RESULT 15			
71571			
AAAT71571 standard; DNA; 4767 BP.			
AC	AAAT71571:		
XX			
DT	11-DEC-2000 (first entry)		
XX			
DE	S. aggregatum PKS cluster ORF8/ORF9 homolog DNA.		
XX			
KW	PKS pathway; polyunsaturated long chain fatty acid; plant; transgenic;		
KW	polyketide-like synthesis; pufa; dietary supplement; intravenous feeding		
KW	manutrition; cooking oil; cooking fat; margarine;		
KW	docosahexenoic acid production; eicosapentenoic acid production; ds.		
XX			
OS	Schizochytrium aggregatum.		
XX			
PN	WO200042195-A2.		
PD	20-JUL-2000.		
XX			
PF	14-JAN-2000; 2000WO-US00956.		
XX			
PR	14-JAN-1999; 99US-0231899.		
XX			
PA	(CALJ ) CALGENE LLC.		
XX			
PI	Facclotti D, Metz JG, Lassner M;		
XX			
	WPI: 2000-476063/41.		
	P-PSDB: AAB10484.		
PT	New DNA sequences encoding for polyketide (PK)-like synthesis pathway		
PT	genes from Shewanella, Vibrio and Schizochytrium; useful for creating		
PT	transgenic plants that express poly-unsaturated long chain fatty acids		
PS	Claim 6; Fig 27C1-2; 302pp; English.		
XX	This invention describes novel DNA sequences encoding for polyketide		
CC	(PK)-like synthesis (PKS-like) pathway genes from Shewanella, Vibrio		
CC	and Schizochytrium. The nucleic acids are useful for isolating related		
CC	molecules or in methods to detect organisms expressing the PKS-like		
CC	genes. They are also useful for creating transgenic plants that express		
CC	poly-unsaturated long chain fatty acids. The poly-unsaturated long chain		
CC	fatty acids produced recombinantly are useful as dietary supplements for		
CC	patients undergoing intravenous feeding or for preventing or treating		
CC	malnutrition. The poly-unsaturated long chain fatty acids can also be		
CC	incorporated into cooking oils, fats or margarine formulated so that in		
CC	normal use the recipient receives a desired amount of poly-unsaturated		
CC	long chain fatty acid. The nucleic acids are also useful in large scale		
CC	production of docosahexenoic acid and eicosapentenoic acid, and for the		
CC	modification of the fatty acid profile of host cells and edible plant		
CC	tissues and/or plant parts. Transgenic production of polyunsaturated		
CC	fatty acids in particular host cells allows quicker purification from		

CC natural sources such as fish or plants. This sequence encodes the  
CC Schizochytrium aggregatum PKS protein cluster ORF8/ORF9 homolog composed  
CC of clones LIB 84-015-D5, LIB 81-042-B9 and a bridging PCR product.  
XX  
SQ Sequence 4767 BP; 922 A; 1764 C; 1267 G; 792 T; 2 other;

Query Match	4.4%	Score 45	DB 21	Length 4767
Best Local Similarity	42.9%	Pred. NO. 0.59		
Matches 222	Conservative 0	Mismatches 295	Indels 0	Gaps 0

QY	335	ccgctccacacatccctcgaagcgcaacccggagctctcgagggagaaagtgtggagacgctgctc	394
Db	3083	ccggccccggccccagacacatccgaacctcgaagacgtcgaagccgagctccctcgactcgatg	3142
QY	395	tcatgtctccctctgagaaatttggtccctcgaaccttgagggttcaccccgcgagaccgtctcacag	454
Db	3143	ccccgctctactctctcgacagaccgcgacggcgccagctctcaagaaagcacacagctggtg	3202
QY	455	ccttggaagaaaggtctcttgagaaaaagggctggttggtccacgcagctctcttcacccaaccg	514
Db	3203	ccctccggccagagccacacatcgtgcagggccctgcagctctcgcgagacctcggttgacccgctctc	3262
QY	515	tcaacagggagacgcgcggcgccatctcgcaacctcttggtggcgtctccggctctggaagccagagaa	574
Db	3263	tcatgtgagacctgaagctgctgctggccccgtgtgacaacggcgcgacatgagccaagggcatgtg	3322
QY	575	ccagagctccaacctctgaaagactaaactaacctcccttgagagaaactcgccctagaacatgtgcc	634
Db	3323	ccctggcgagaaactctgtcatccgcgcgcgcgaacagctgcagaagctctccctgtctctgttcggc	3382
QY	635	acggcagctgctctctcttcacacttggttcgaaggccctacacaggaacacagaaatccgcgccaccc	694
Db	3383	ggcgccctcccccattgcaacacagctgctggcgccgctctcgagagaatctcaagccgcgcctgcctc	3442
QY	655	tggaacgtgctggtctcctcttggtgtcagggaggaaggtgtatgcccaacacagagacgtgtg	754
Db	3443	agggtcccccctaaagccgttcaaacctcatccatccgcgtcttttcagacgaacactcggaaaggcca	3502
QY	755	ccgtcttgaaacctggccatctctgtctccctccggagcttggaagcgttccaaacctccggcgcc	814
Db	3503	acgtctgactctctctctctcgaaagagggcgtctacgtgtgtgtgaagccctccgcattcatagacc	3502
QY	815	ttatcacacctcatgtcagctacatctgcgactcctc	851
Db	3563	tcaccccgcagctgtgtctgactacggcgccgcgctc	3599

Search completed: May 29, 2002, 10:21:39  
Job time: 4291 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2002, 08:12:58 ; Search time 1687.79 Seconds

(without alignments)  
8204.739 Million cell updates/sec

Title: US-09-664-186-4  
Perfect score: 1026  
Sequence: 1 gtgaagaacgaacaaacctt.....ccctcaagagagcctcctga 1026

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estbta:\*  
2: em\_estbhum:\*  
3: em\_estlin:\*  
4: em\_estlmu:\*  
5: em\_estrov:\*  
6: em\_estropl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	54.6	5.3	925	12	CNS0091P
2	51.8	5.0	925	12	CNS0091P
3	50.6	4.9	637	10	BI960118
4	49.2	4.8	502	10	BM335757
5	49.2	4.8	759	10	BF259495
6	48	4.7	524	9	AV436476
7	47.6	4.6	463	10	BG946638
8	47.6	4.6	539	10	BM332294
C 9	47.6	4.6	564	10	BM075274
10	47	4.5	1101	12	CNS01757
11	46.6	4.5	497	9	AA388278
12	46.4	4.5	591	10	BI960325
13	46.2	4.5	513	9	AM927362
14	46.2	4.5	534	9	AM331023
15	46.2	4.5	548	9	AM927363
16	46	4.5	490	9	AM520296
C 17	46	4.5	545	10	BE511221

18	46	4.5	588	9	AM566229	AM566229
19	46	4.5	712	10	BM351809	BM351809
20	45.8	4.5	501	10	BI273373	BI273373
21	45.8	4.5	562	10	BF484304	BF484304
22	45.6	4.5	762	10	BF265677	BF265677
23	45.6	4.4	526	12	BH219902	BH219902
24	45.6	4.4	530	12	BH219884	BH219884
25	45.4	4.4	560	10	BF478389	BF478389
26	45.4	4.4	676	10	BG905572	BG905572
27	44.8	4.4	376	10	BE918312	BE918312
28	44.8	4.4	664	10	BM328060	BM328060
29	44.6	4.3	342	9	AM564526	AM564526
30	44.6	4.3	358	10	BG273412	BG273412
31	44.6	4.3	388	10	BF656946	BF656946
32	44.6	4.3	412	9	AM564034	AM564034
33	44.6	4.3	415	10	BE596249	BE596249
34	44.6	4.3	427	10	BI074677	BI074677
35	44.6	4.3	468	10	BG240338	BG240338
36	44.6	4.3	469	10	BE367548	BE367548
37	44.6	4.3	478	10	BG240581	BG240581
38	44.6	4.3	493	10	BG946964	BG946964
39	44.6	4.3	505	10	BE365880	BE365880
40	44.6	4.3	525	10	BE594368	BE594368
41	44.6	4.3	535	10	BE360673	BE360673
42	44.6	4.3	546	9	AM671463	AM671463
43	44.6	4.3	558	10	BE726101	BE726101
44	44.6	4.3	598	10	BI716428	BI716428
45	44.6	4.3	598	10	BE596557	BE596557

#### ALIGNMENTS

RESULT 1  
CNS0091P/c  
LOCUS  
DEFINITION  
Drosophila melanogaster genome survey sequence Tm3 end of BAC #  
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION  
AL053013  
VERSION  
AL053013.1 GI:4934461  
KEYWORDS  
SOURCE  
ORGANISM  
Drosophila melanogaster  
fruit fly.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mammosser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp. The same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

#### FEATURES

SOURCE  
1..925  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"

BASE COUNT	ORIGIN	120 a	61 c	61 g	172 t	511 others
/clone="BACR19D16" /note="end : TET3"						

Query Match	5.38;	Score 54.6;	DB 12;	Length 925;
Best Local Similarity	15.88;	Pred. No. 0.099;		
Matches 63;	Conservative 163;	Mismatches 172;	Indels 0;	Gaps 0.

[illegible]

LOCUS	DEFINITION	FEATURES
CNS00091P	925 bp DNA	linear
CNS00091P	Drosophila melanogaster genome survey sequence TE13 end of BAC # BAC119016 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	GSS 03-JUN-1999

ACCESSION	AL053013	GI:4934461
VERSION	AL053013.1	
KEYWORDS	GSS.	
ORF	fruit fly.	
ORGANISM	Drosophila melanogaster	

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

**COMMENT**

Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage  
BP 191 91006 Evry-CorDEX - FRANCE (E-mail: seqref@genoscope.cns.fr  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))  
determination of this Berkeley sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see: <http://www.fruitfly.org/TheBDGP/Drosophila>  
melanogaster BAC library was prepared by Kazuo-ya Oosagawa and  
Aaron Mammosser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPc1-98 and was constructed by partial  
restriction digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be

```

FEATURES
source      found at http://bacpac.med.buffalo.edu/drosophila_bac.htm
            Location/Qualifiers
            1..925
            /covariation="Drosophila melanogaster"

```

BASE COUNT	120 a	61 c	61 g	172 t	511 others
ORIGIN					

Query Match	5.0%;	Score 51.8;	DB 12;	Length 925;
Best Local Similarity	13.7%;	Pred. No. 0.39;		
Matches	54;	Conservative 172;	Mismatches 167;	Indels 0;
				Gaps 0;

[illegible]

RESULT	3
B1960118	
LOCUS	B1960118
DEFINITION	B1960118 637 bp mRNA linear EST 22-OCT-2001 HVSNErn0023E18f Hordeum vulgare rachis EST library HVCdNA0015 (normal) Hordeum vulgare cDNA clone HVSNErn0023E18f, mRNA sequence.

ACCESSION	BY560118	
VERSION	BT560118.1	GI:16311373
KEYWORDS	EST.	
SOURCE	barley.	
ORGANISM	Hordeum vulgare	

## REFERENCE

## AUTHORS

**TITLE**

JOURNAL

**COMMENT**

J. V. Yates, K. and Malm, D.  
Development of a genetically and physically anchored EST resource  
for barley genomics: Morex rachis cDNA library  
Unpublished (2001)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: twing@clemson.edu  
Total hg bases = 579



Seq primer: AATTAACCTCCTAAGGCG  
High quality sequence start: 5  
High quality sequence stop: 633.  
Location/Qualifiers

# FEATURES

Source

1. 637  
/organism="Hordeum vulgare"  
/cultivar="Morex"  
/db\_xref="taxon:4513"  
/clone="HVSME0023E18f"  
/clone\_lib="Hordeum vulgare rachis EST library HVCNDA0015 (normal)"  
/tissue\_type="rachis"  
/lab\_host="TJC121"  
/note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2: XhoI. Plants were grown at Washington State University, Pullman, WA in a greenhouse, the rachises were excised and frozen in liquid nitrogen (Kleinholz lab). In the TJC Close lab at the University of California, Riverside total RNA was prepared, poly(A) was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see  
<http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinholz A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/g9pages/dgn/31/cover.html>)"

BASE COUNT 125 a 249 c 158 g 105 t  
ORIGIN

Query Match 4.9%; Score 50.6; DB 10; Length 637;  
Best Local Similarity 46.3%; Pred. No. 0.67;  
Matches 167; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

447 cgtccacgctggaagaagctccttgaagaagaagccttggtccacgacgtcctca 506  
133 CATCAACCGGAGAACTACGAGCTGGGGCTGCCGTGCATGAGAAGCCCGCGTGGCGCA 192  
507 ccaaacgctcaacgagcgccgagccatcgaccccttggcgctgagcgtgagcc 566  
193 CAGATCGACTTCGGGAGGGCCGGCGCTCCCGTCTGAGAGCCCTCTCGAGAGCA 252  
567 agggaaagcagagctcaacctgaactacatcccttggaagaacctgcacctaga 626  
253 GGCCAAACGAGGACCTTCACCTTCGCTTCGTGGAGCGGACGAACAAGCAACTACCTCAA 312  
627 catgagcaacgagcgctgtctcccttcaactggttcaagaagctaccagagcaacgagatccg 686  
313 CTACACGAGACGGCTCATGAACTCGTCAAGAGTGGGGCTCTCTCGCTACGCAACAC 372  
687 ccccaaccttgagcgtgtgttcttgcgtcaggggaaaggtgtagtcccaacaccaa 746  
373 CCTCTGGAAGGGCTCCGTCGTGCTCCCGCGAGGCCCCCATGGCAAGTACATCCGCTA 432  
747 gaccgtgacgttgacttgagctgagctcattcgttctcccgaggtgtagcgcttccaact 806  
433 CTACCGGACTTGGTCTCTCACTCAACAGAGGCCCTCGCGACCAAGCGCGTGAAGAT 492  
807 c 807  
493 C 493

RESULT 4  
BM335757/c  
LOCUS  
DEFINITION  
502 bp mRNA linear EST 16-JAN-2002  
MEST160-F05.T3 ISUM5-RN Zea mays cDNA clone MEST160-F05 3', mRNA

ACCESSION  
BM335757  
VERSION  
BM335757.1 GI:18165918  
KEYWORDS  
EST.  
SOURCE  
Zea mays.  
Zea mays.  
Zea mays.

REFERENCE  
AUTHORS  
TITLE  
Men,T.J., Qiu,F., Guo,L., Ashlock,D.A and Schnable,P.S.  
1 (bases 1 to 502)  
Expressed Sequence Tags from B73 Maize: various stages and tissues including seedlings treated with a variety of hormones

JOURNAL  
COMMENT  
Unpublished (2001)  
Contact: Patrick S. Schnable  
Schnable Laboratory  
Iowa State University  
405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA  
Tel: 515-294-0975  
Fax: 515-294-2299  
Email: [schnable@iastate.edu](mailto:schnable@iastate.edu)  
Individual basecall and confidence value were assigned using the phred software.

(<http://depts.washington.edu/ventures/collabtr/direct/index.htm#b> rt). Overall sequence quality assessment and vector trimming were conducted using the Lucy software (<http://www.tigr.org/softlab/lucy>). Lucy parameters were set to ensure an overall trimmed quality of 97.5% or better without any vector fragments in the chosen high-quality region of each sequence. Low-quality bases between the poly-T and the high-quality region were replaced with N's to serve as spacers.

PCR Primers  
FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)  
BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)  
Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).  
Location/Qualifiers

## FEATURES

source

1. 502  
/organism="Zea mays"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="MEST160-F05"  
/clone\_lib="ISUM5-RN"  
/tissue\_type="mixed"  
/lab\_host="DH10B"  
/note="Vector: pT73PAC; Site.1: EcoRI; Site.2: NotI. Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG), Mixed mature tissues (17, 21, 38, 69, 77 DAG), kernels (3, 5, 10, 15, 20, 25, 30, DAG), Adventitious roots (65 DAG), Tassel (3-39 cm, 53 and 56 DAG), Immature ear (0.2-3.0 cm, 33, 36, 59 DAG), Husk (73 DAG), Silk, unpollinated first ear, ear shank, etiolated seedlings, callus, Cytokinin-treated callus, Anaerobic treated seedlings, NAA (a-Naphthalene acetic acid)-treated seedlings, Kinetin-treated seedlings, ACPG (1-aminocyclopropane-1-carboxylic acid)-treated seedlings, Brassinolide-treated seedlings, ABA (Abscisic acid)-treated seedlings, GA (Gibberellic acid)-treated seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA molecules were generated as follows. First-strand cDNA was prepared from oligo-dT selected mRNA by priming with a NotI oligo-dT primer (5', AACTGGAAGATTGGCGCCGACGAGATTGTTTGTGTTT). The resulting DNA:RNA hybrid was treated with RNase H and used as a template for DNA PolI-catalyzed second strand synthesis. After the addition of EcoRI adaptors, the ds-cDNAs were digested with NotI and size-selected. The resulting molecules were directionally cloned into the EcoRI and NotI sites of the pT73PAC vector. The library then went through one round of normalization to COT value



TITLE	Generation of 10,154 expressed sequence tags from a leaky gametophyte of a marine red alga, <i>Porphyra yezoensis</i>
JOURNAL	DNA Res. 7, 223-227 (2000)
MEDLINE	20363100
COMMENT	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES	
source	1. 524 /organism="Porphyra yezoensis" /strain="TU-1" /db.xref="taxon:2788" /clone="PS002f12_r" /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT	91 a 176 c 196 g 61 t
ORIGIN	
Query Match	4.7%; Score 48; DB 9; Length 524;
Best Local Similarity	45.6%; Pred. NO. 2.3;
Matches 208; Conservative	0; Mismatches 245; Indels 3; Gaps 1;
Oy 284	aagcgtcacaagcagcgagcgtcaagaatcttcgaacctctccaaagagatgccgctcca 343
Db 69	AGCGCGCGGAGCGCGCTCGCGGCTGGCGGCGCTGACTGGAAGACAGACTCGCGGCTCA 128
Oy 344	ccatccctcacaagcaaccgagagctctcyagaagatgagagagcagcgtcgtctcatgtctc 403
Db 129	CCCCCGCGAGCGCAAGCGCGCGCAGCAGCGGCTGTGAGAGAGCGCGGCTTTGAAAAAG 188
Oy 404	cccttgaaatgttggcctcaacctcgggggttaccocggcagaacctccaagctctgaaga 463
Db 189	CCAAAGAGATGTTTTCACACCGCCCGGCTGGCGGCGCGCAGCGCGCGCGGCGGCTGACCG 248
Oy 464	aggtccttgagaaaaagggcctgtgtgycacacgacgtcccttaccaccaaacgctcaacgg 523
Db 249	ACTCATCGACACCTTCCCGCCCGCCCGCTGTCGAGCGGAGCCAGCCGCGGCTTTGGCGGCG 308
Oy 524	agcgcggagcc--atcgacaccttctgggcccgtccgctgtgagccaggaagcagcagc 580
Db 309	TGCTGGGCGCCCAAGTGTGGCACCGCTCTGGTGTCCGCGCGGCGGCGCTTCAAGTACGTGGGGA 368
Oy 581	tcaacctgagacactacatctaccccttgaggaagacgtgcgcttagacatgacgcacaagcgc 640
Db 369	TGCTCAAGGGCATCTCCGCAACCGCCGTCGCGGACTGTGCGCAGACGACGTCAAGGAGAC 428
Oy 641	tgtctccttcaacttgygtlcaagagcttacaagacacgaagaaatccgcccacaccttgagc 700
Db 429	TGCTGTGCGGTGTGCACAGCGTGTGGGAATGAGAAACTCAAGAGAGAGGCTTCAAGCGGAGCA 488
Oy 701	tgtgtgtcctctggtgctcagggagaaaggtgattgc 736
Db 489	AGGGGAGAGAAACCGGTGAAGAGAGTCTGC 524
RESULT 7	
LOCUS	BG946638/c 463 bp mRNA linear EST 12-JUN-2000
DEFINITION	949012011.X1 949 - Juvenile leaf and shoot cDNA from Steve Moose
ACCESSION	BG946638
VERSION	BG946638.1 GI:14364827
KEYWORDS	EST.
SOURCE	Zea mays.
ORGANISM	Zea mays. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE	1 (bases 1 to 463)

**AUTHORS** Walbot, V.  
**TITLE** Maize ESTs from various cDNA libraries sequenced at Stanford University  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave,  
Palo Alto, CA 94304, USA  
Tel.: 650 723 2227  
Fax: 650 723 8221  
Email: walbot@stanford.edu  
Plate: 949012 row: C column: 11.

**FEATURES**  
source  
1..463  
/organism="Zea mays"  
/cultivar="W64A"  
/db\_xref="taxon:4577"  
/clone\_id="949 - Juvenile leaf and shoot cDNA from Steve Moose"  
/tissue\_type="Immature leaf primordium and vegetative meristem"

/dex\_stage="4 stages from 3-13 days after imbibing"  
/lab\_host="E. coli XL0LR"  
/note="Organ: juvenile vegetative shoots; Vector: pAD-GAL4-2.1; Site\_1: EcoRI; Site\_2: XhoI; Equal amounts of total RNA by weight from 4 tissue sources (see below) were pooled, polyA+ RNA isolated, and cDNA synthesized for EcoRI (5') and XhoI (3') directional cloning into lambda Hybridzap vector from Stratagene. Tissue Sources: 1. Whole shoots 3 days after sowing/imbibing in wet soil. 2. Basal 1.5 cm shoots 6 days after sowing - includes yellow portions of developing leaves 1-5, primordia from 6-8, and the vegetative apex. 3. Non-green portions of developing leaves 4-5 and the vegetative apex, including adult leaf primordia, 9 days after sowing. 4. Partially expanded and greening leaves 4-5 at 13 days after sowing."

**BASE COUNT** 79 a 132 c 175 g 77 t

**ORIGIN**

```

Query Match          4.6%; Score 47.6; DB 10; Length 463;
Best Local Similarity 48.9%; Pred. No. 2.8;
Matches 128; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
```

OY	344	ccatcccctacgcaaccggagagctctggaagggtgggaacggtcgcttcattgctc	403
Dd	358	CCAACCCCTACGACCAGCGCCTCCCACTGTGGAGATGCATCACGCCGTCGCGCG	299
OY	404	cacctgagatgttggccccccaactcgggggtcacccggacaagccgctcaagcgttgaa	463
Dd	298	GCAAGGAGGTGGCTTCGGGACCATGCGGACCCGGCTGGATTCCCAGCGCTTCCA	239
OY	464	aggtctctgaagaaaagggcctggtggtgcacacgcatctaccacaacacggtcaacgggg	523
Dd	238	CGGCGCTGGAGATCCCCCGGCAGAGTGCCCTACGACTTCCTGCTTCCCTCGTCAGGAGCG	179
OY	524	agagcggggccatcggcaacctttggtggccgttcggctgaagcccaaggaaagcaggtcca	583
Dd	178	TGCGGCGGAGACGGGACAATCGACTACGAGACCTCAGGTGGGCTCACCGCTGCACCTCCCA	119
OY	584	ccctgagcagctacatctacc	605
Dd	118	TGCTGGCGCACTTCACCATCC	97

**RESULT** 8  
BM332294  
LOCUS BM332294/C 539 bp mRNA linear EST 16-JAN-2002  
DEFINITION MEST154-H06.t3 ISUM5-RN zea mays cDNA clone MEST154-H06 3', mRNA sequence.  
ACCESSION BM332294  
VERSION BM332294.1 GI:18162455  
KEYWORDS EST.















**THIS PAGE BLANK (USPTO)**